Building Up Research Capacity at Minority Institutions

Commissioned Paper for the Committee on the Development of a Plan to Promote Defense Research at Historically Black Colleges and Universities, Tribal Colleges and Universities, Hispanic-Serving Institutions, and Other Minority-Serving Institutions

Ruby Zhang
March 2024
# Contents

1 Introduction..................................................................................................................... 3  
  INSTITUTION CLASSIFICATION ............................................................................. 3  
  FRAMEWORK ........................................................................................................... 5  
  FIGURE AND TABLE ATTRIBUTES ...................................................................... 7  
2 Data Sources.................................................................................................................... 8  
  CARNEGIE CLASSIFICATION ........................................................................... 8  
  IPEDS ....................................................................................................................... 9  
  NSF: HERD AND S&E RESEARCH FACILITIES SURVEYS................................. 9  
  KEY TAKEAWAYS ............................................................................................... 11  
3 Carnegie Classifications................................................................................................. 13  
  CARNEGIE CLASSIFICATIONS OVER TIME ...................................................... 13  
  NEW CARNEGIE CLASSIFICATION OF 2025 ....................................................... 14  
  KEY TAKEAWAYS ............................................................................................... 15  
4 How Far are MIs from R1s and R2s? ............................................................................ 17  
5 R&D Expenditure.......................................................................................................... 19  
  R&D EXPENDITURE DISTRIBUTIONS ................................................................ 19  
  R&D EXPENDITURE BY FIELD IN 2021 ............................................................. 20  
  CAPITALIZED EQUIPMENT EXPENDITURE BY FIELD IN 2021 ...................... 23  
  KEY TAKEAWAYS ............................................................................................... 24  
6 Facility Space.................................................................................................................. 26  
  FACILITY SPACE DISTRIBUTIONS ...................................................................... 26  
  FACILITY SPACE BY FIELD IN 2021 ................................................................. 27  
  FACILITY COST BY FIELD IN 2021 ................................................................. 30  
  KEY TAKEAWAYS ............................................................................................... 32  
7 R&D Expenditure versus Facility Space....................................................................... 34  
  QUANTILE REGRESSION OF R&D EXPENDITURE ON FACILITY SPACE .......... 37  
  KEY TAKEAWAYS ............................................................................................... 39  
8 R&D Expenditure and Facility Space Over Time ......................................................... 40  
  R&D EXPENDITURE FROM 2012 TO 2021 ........................................................... 40  
  FACILITY SPACE AND COSTS FROM 2011 TO 2021......................................... 43  
  KEY TAKEAWAYS ............................................................................................... 44  
9 Conclusion...................................................................................................................... 47  
  CARNEGIE CLASSIFICATIONS ........................................................................... 47  
  INVESTMENT BY FIELD AND INSTITUTION TYPE........................................... 48  
  NEXT STEPS ........................................................................................................... 50  
References............................................................................................................................ 52
1

Introduction

This paper aims to characterize the current research capacity of Minority Institutions (MI) and explore potential investment avenues, using survey data from the National Science Foundation (NSF). The paper compares MI to institutions with very high research activity (R1) and high research activity (R2) as classified by the Carnegie Classification of Institutions of Higher Education (CCIHE). The analysis examines where MI currently stand in the Carnegie Classification, and how they compare to R1s/R2s in terms of research and development (R&D) expenditure and facility space broken down by field.\(^1\) However, due to left truncation in the data, while R1s/R2s are well represented, the selection focuses on the highest-spending MI. Despite the missing data problems, only by first understanding the current state of research using the best available public data can forward-looking policy suggestions then be made. To that end, it is important to keep in mind and understand what the institution objectives are, and under what framework institutions operate. This paper first describes the institution types, to inform readers of the different objectives that institutions may have.

INSTITUTION CLASSIFICATION

Throughout its entirety, the paper uses the following six institution classifications:

- **HBCU**: Historically Black Colleges and Universities
- **HSI**: Hispanic-Serving Institutions
- **TCU**: Tribal Colleges and Universities
- **R1**: Very high research activity
- **R2**: High research activity
- **R3**: Doctoral/professional universities

MI encompasses HBCUs, HSIs, and TCUs. Other includes R3s if it is not a separate category in the table

\(^1\) The focus is on science and engineering fields, including psychology and social sciences.
or figure in which it appears. MI designations are based on the Integrated Postsecondary Education Data System (IPEDS), and research institution designations are based on the 2021 CCIHE. Note that the Carnegie Classification system excludes TCUs and Special Focus Institutions. The latter are institutions that offer academic programs concentrated in a specific set of fields, such as health, arts, law, theology, engineering, business, and other professions. Special Focus Institutions are excluded from the analysis, as they are not given a Carnegie Classification. Medical schools and research facilities (e.g., Van Andel Institute) are also excluded from the analysis.

There are two types of designations for the three categories of MIs in this paper: historical (HBCUs and TCUs) and threshold based (HSIs) and are described next.

HSIs are defined by the Hispanic Association of Colleges and Universities (HACU) as “not-for-profit institutions of higher learning with a full-time equivalent undergraduate enrollment that is at least 25% Hispanic” (HACU, 2024). This definition is in Title V of the Higher Education Act (U.S. Department of Education, 2024a). HACU maintains a list of HSIs that is updated annually based on the enrollment numbers of the previous year. Thus, the set of HSIs is subject to change over time, since it is based on student composition thresholds. The type of institutions are also different from HBCUs and TCUs. For example, HSIs include many large, R1 state schools in the University of California (UC) and University of Texas (UT) systems, which have high local Hispanic populations. The subsequent analysis highlights HSIs as one of the most heterogeneous institution types in terms of research capacity due to the large variation in size and research intensity.

HBCUs are defined by the Higher Education Act of 1965 as “any historically black college or university that was established prior to 1964, whose principal mission was, and is, the education of black Americans” (U.S. Department of Education, 2024c). Unlike HSIs, the list of HBCUs is not fluid over time, as it is based on historical designation/mission. There are currently 102 accredited HBCUs.

TCUs are self-determined and chartered by each respective tribal government, constituting a total of 35 accredited institutions serving students as of this paper (American Indian Higher Education Consortium, 2024). Along with improving educational and career opportunities for Native American students, TCUs also “maintain, preserve, and restore Native languages and cultural traditions . . . and often serve as anchors in some of the country’s poorest and most remote areas” according to Executive Order 13592 (U.S. Department of Education, 2024b).
While the number of accredited TCUs has increased over time, the institution mission is not based on a threshold of student composition but rather also on preserving Native American culture and serving the community.

**FRAMEWORK**

In an economic framework, each institution wishes to maximize an *objective function*, subject to budget constraints. The objective function is the goal of the institution, and the budget constraints are the limitations on funding and resources. There are several ways that the goal of “improving research capacity” can be formulated, each with different policy implications. Using R1s as the benchmark, it becomes clear that there are fields that require massive amounts of capital investment and funding (e.g., life sciences, engineering, physical sciences) and fields that use far less capital and total R&D spending (e.g., mathematics and statistics, psychology, social sciences). Thus, the first question to keep in mind when reading the paper is what objective function a particular institution has:

1. Does an institution aim to increase research capacity across *all* fields and follow in the path of an R1 institution? Then fields with large disparities in R&D spending suggest potential areas for investment. A related model is whether an institution wishes to reach a minimum research level across all fields as opposed to maximizing all fields.
2. Does an institution aim to become good at certain fields and specialize? Fields of specialization may be due to historical missions of the institution, local industry, or local community needs. Alternatively, institutions may wish to specialize in fields that require less capital and are more feasible to build up.
3. Does an institution wish to increase net research capacity or focus on increasing research capacity per student? Depending on needs of the local community and the population that the institution is serving, the objective may be to increase research capacity per student as opposed to maximizing the total research output via increasing the number of students.

The next common question in an economics framework is the mode of competition that institutions face, taking “research output” as the industry. There are both inter- and intra-
institution considerations:

1. To what extent do institutions compete or collaborate with each other? For example, institutions may collaborate on research projects and share resources, or they may compete for the same pool of funding, researchers, students, and so forth. In the latter case, having an R1 institution may be beneficial for knowledge spillovers, but in the former case, it may also lead to brain drain. A related question is the peer group of a given institution. Are HSIs competing with R1s and R2s for research funding? Is this true for TCU's and HBCUs?

2. To what extent are fields within institutions independent agents? Firms may operate as a multiproduct firm trying to maximize an overall objective across fields, but it is also possible that fields within an institution are competing for resources and can be thought of as separate industries. The latter would complicate the objective function of the institution, going back to the first set of questions.

The paper does not answer these questions, but it provides the data and analysis that can inform the answers to these questions. Apart from cross-sectional frameworks, research funding is also often a dynamic process and suffers from the Matthew effect (i.e., the rich get richer), which can influence policy decisions aimed at improving long-term research capacity (Qiu, 2023). This paper also does not have data on eventual downstream outcomes of research spending, such as patents, publications, or citations. However, there are many papers that show that funding leads to real outcomes (Azoulay et al., 2018; Tabakovic and Wollmann, 2019). Different institution types have different objectives and modes of competition, which is important to contextualize the paper’s findings.
FIGURE AND TABLE ATTRIBUTES

All figures and tables are each in their own document to simplify viewing, with the intention of being read side-by-side with the main report. The figures and tables are numbered according to the order they appear in the report. Special Focus Institutions, medical schools, and research facilities are excluded from all analysis.

All bar graphs with numbers on top of the bars represent the number of institutions that have an entry in the Higher Education Research and Development (HERD) Survey or Survey of Science and Engineering (S&E) Research Facilities for that field. Note that not all institutions have data for all fields even if they participate in the survey. The numbers do not represent the actual value being graphed. Tables reporting the measure (e.g., R&D expenditure or facility space) per undergraduate or graduate student are based on total enrollment as of fall 2020.

The rest of the paper is organized as follows. Section 2 describes all the data sources used for analysis, consisting of the CCIHE, IPEDS, and the NSF HERD and S&E Research Facilities surveys. Section 3 describes the historical and future Carnegie Classification of Institutions. Section 4 briefly examines the geographic distance between MIs and R1s/R2s. Section 5 compares R&D expenditure across institution types and fields. Section 6 compares facility space across institution types and fields. Matching R&D expenditure and facility space data, Section 7 correlates spending and space controlling for student body size. All analysis mentioned up to now uses the latest cross-sectional data from 2021; thus, Section 8 briefly examines R&D expenditure and facility space trends by field over the last decade. Finally, Section 9 concludes the paper.
2

Data Sources

CARNEGIE CLASSIFICATION

This paper uses the 2021 Carnegie Classification of Institutions of Higher Education as the base file for the universe of postsecondary institutions because the goal of this paper is to compare the research capacity of minority institutions to that of institutions classified as R1 (very high research activity) and R2 (high research activity) under the Carnegie system (American Council on Education, 2021a). Section 3 describes the current classification criteria. Developed by the American Council on Education, the CCIHE has been the leading framework in categorizing U.S. institutional diversity in postsecondary education. The CCIHE is based on data from the Integrated Postsecondary Education Data System. The classification is updated every 3 years, with 2021 being the most recent version. From the 2021 public data file, there are the classifications for all institutions in 2000, 2005, 2010, 2015, 2018, and 2021. There are also indicator variables for HBCUs and TCUs (from IPEDS), an indicator variable for HSIs (from HACU), as well as undergraduate and graduate enrollment in fall 2020 (from IPEDS) that are used in this paper.
IPEDS

This paper uses Integrated Postsecondary Education Data System data indirectly via the public data file provided by the CCIHE. IPEDS is the primary source for data on postsecondary institutions in the United States and is collected by the National Center for Education Statistics. The paper utilizes location variables (county and longitude/latitude) directly from the 2022 IPEDS public data file (National Center for Education Statistics, 2024).

NSF: HERD AND S&E RESEARCH FACILITIES SURVEYS

The primary data sources of institutional R&D activity are the Higher Education Research and Development Survey and the Survey of Science and Engineering Research Facilities. Both surveys are conducted by the National Science Foundation. The annual HERD survey collects information on R&D expenditures at U.S. colleges and universities broken down by field and expenditure type. The biennial S&E Research Facilities survey collects information on the amount of space and costs for research and development facilities, also broken down by field. The HERD and S&E Research Facilities surveys collect data on all U.S. academic institutions reporting at least $150,000 (NSF, 2022a) and $1 million (NSF, 2022b) in R&D expenditures in the previous fiscal year, respectively.

This paper uses the field delineation present in each survey. While the two surveys share mostly the same field delineation, the S&E Research Facilities survey has a finer breakdown: life sciences is separated into biological sciences, health sciences, agricultural sciences, and natural resources. Psychology and social sciences are part of science and engineering in the NSF surveys, so non-S&E exclude psychology and social sciences. R&R is an abbreviation for repair and renovation in the S&E Research Facilities survey.

The focus of this paper is on research costs and facility space as opposed to the student body. There is no analysis on the undergraduate or graduate student population in this version of the paper. Total undergraduate and graduate enrollment in fall 2020 is only used insofar as a weighting mechanism to adjust for spending or facility space per student.

---

2 Agricultural sciences and natural resources are two categories in the facilities survey post-2015.
Institution Counts

Since there is an R&D expenditure cutoff to be in the NSF surveys, there is selection on the largest R&D institutions. Table 1 shows the number of institutions in the HERD and S&E Research Facilities surveys by institution type. There were 101 HBCUs, 332 HSIs, and 35 TCUs in the country in 2021. However, only 55 HBCUs, 81 HSIs, and 5 TCUs are present in the HERD survey. The S&E Research Facilities survey has even fewer institutions, with only 40 HBCUs, 26 HSIs, and 1 TCU. Approximately 40 to 50 percent of HBCUs are in the S&E Research Facilities and HERD surveys, respectively. However, less than a quarter of HSIs are present in the HERD survey and around 8 percent are present in the S&E Research Facilities survey. TCUs are highly underrepresented in both surveys, with only one TCU present in the S&E Research Facilities survey. Thus, all subsequent analysis represents a skewed picture of MI research capacity (only the highest-spending institutions are present).

On the other hand, the HERD survey collects data on the universe of R1 and R2 institutions, and less than half of R3s. However, the S&E Research Facilities survey unexpectedly does not cover all R1 and R2 institutions. Note the abundance of institutions outside of MI and Carnegie classifications. Despite the importance of research facilities in building research capacity, the NSF data does not provide a complete picture even for R1s and R2s.

Institution Counts by Field

While many institutions do not have a presence in all the fields of the NSF surveys, this is reported differently in the HERD and S&E Research Facilities surveys. The HERD survey is unbalanced, with institutions reporting R&D expenditures only for fields that they have. Note that this includes reporting zero R&D expenditure for some fields. However, the S&E Research Facilities survey is balanced, with institutions reporting zero facility space for fields that they do not have. This is important to keep in mind when comparing R&D expenditure and facility space across fields. Across field comparisons within institution type is not necessarily balanced for the HERD survey. For all figures, the number of institutions reported is stated for every field.

Table 2 shows the number of institutions in the HERD and S&E Research Facilities
surveys by field. The first column counts the number of institutions in both the HERD and Facilities surveys, the second column counts the number of institutions only in the HERD survey, the third column counts the number of institutions only in the S&E Research Facilities survey, and the last column is the total number of institutions in either the HERD or Facilities survey for that field. For nearly all fields, the number of institutions that are only in the HERD survey exceeds the number of institutions in both the HERD and the Facilities surveys. For the most part, the S&E Research Facilities survey is a subset of the HERD survey, as expected. However, the subsetting is not perfect either. This distinction is crucial for the analysis in Section 7, where R&D expenditure is compared to facility space across various fields. For reference, Tables 32 and 33 list all HBCUs and HSIs that are in the HERD and S&E Research Facilities surveys.³

KEY TAKEAWAYS

The paper draws from a comprehensive set of data sources: the 2021 Carnegie Classifications (supplemented with location data from IPEDS) serves as the universe of postsecondary institutions, while the NSF HERD and S&E Research Facilities surveys provide insight into institutional research capacity by field. A pivotal aspect of this analysis is the utilization of the CCIHE to categorize institutions by their research capacity and MI status. The paper enhances this comparison by incorporating IPEDS data for additional context on the geographical locations and sizes of these institutions.

There is a noticeable underrepresentation of MIs in the HERD and S&E Research Facilities surveys conducted by the NSF, which suggests that the depiction of MI research capacities is skewed, representing only the highest spending institutions within this category. Approximately half of HBCUs are represented, with less than half of HSIs, and nearly no TCUs in the surveys (five in HERD and one in the Facilities survey). In contrast, the HERD survey captures an almost complete snapshot of R1s/R2s’ R&D expenditures, although the S&E

³ Note that any institution with at least one field reported counts as being part of the survey. By this metric, all institutions in the S&E Facilities Research survey are in the HERD survey; thus, the tables list all institutions in the HERD survey (including Special Focus Institutions). The “Both” column indicates that the institution is also in the Facilities survey.
Research Facilities survey does not cover all institutions in these classifications.

Furthermore, the paper sheds light on the heterogeneity in institution participation across different fields in these surveys, complicating direct comparisons of research capacities across institution types and fields. While the HERD survey reports R&D expenditures for fields in which institutions are active, including reporting zeroes, the S&E Research Facilities survey maintains a balanced panel by reporting zero facility space for fields where institutions have no presence. There is thus imperfect overlap between the HERD and S&E Research Facilities surveys (with the HERD survey reporting more than double the amount of institutions in the Facilities survey). Despite these challenges, the paper analyzes research capacity at MIs versus R1s/R2s, with the understanding that the data represents the right tail of research capacity at minority institutions.
3

Carnegie Classifications

CARNEGIE CLASSIFICATIONS OVER TIME

Table 3 provides an overview of how many HBCUs and HSIs are classified as R1, R2, and R3 over time. In 2005, a new classification system was introduced, remaining relatively constant up to 2021. For instance, under these new criteria, Howard University experienced a demotion from R1 to R2 in 2005. Since then, the number of HBCU R2s and R3s have remained relatively constant at four to five until 2018, when most of the same institutions were promoted to R2. This can be seen in Figure 2, which shows the progression of all HBCUs classified as R3 and above since 2005. For example, four institutions (Clark Atlanta, Howard, Jackson State, and North Carolina A&T State Universities) have been R2 institutions since 2005 until 2021, whereas institutions such as Alabama State and Winston-Salem State Universities entered the Carnegie Classification for the first time as R3 institutions in 2021. By 2021, there was a total of 16 HBCUs in the CCIHE (Delaware State University entered as an R3 in 2018, then was demoted in the 2021 classification), which is double the number in 2005.

Table 3 shows steady increases in the number of R1, R2, and R3 HSIs over time since 2005 to a total of 41 institutions in 2021. Note that the new 2005 Carnegie Classification downgraded five R1s and two R2s for HSIs. Figure 3 shows the five HSIs (Texas A&M University, UC Riverside, UC Santa Barbara, UC Santa Cruz, and the University of New Mexico) that have been R1 institutions from 2005 until 2021, with six more HSIs joining the R1 designation over time. Since HSIs are based on student composition thresholds as opposed to historical designation, there are many large state schools with robust research funding such as the UC or UT systems. This leads to very large heterogeneity in the institutions that HSIs comprise.

4 Recall that TCUs are not part of the Carnegie Classification system.
NEW CARNEGIE CLASSIFICATION OF 2025

The latest version of the Carnegie classification defined schools as R3 if they awarded at least 20 research/scholarship doctorates or 30 professional practice doctorates across two programs. Schools are classified as R2 if they spent more than $5 million in total research expenditures and awarded at least 20 research/scholarship doctorates. R1 institutions are defined via an aggregated index comprising research spending in certain fields, number of doctorates, as well as research staff. There is a cap on the number of R1 institutions (American Council on Education, 2021b).

With the current classification, it is hard for institutions to know how to improve in order to progress from R2 to R1. In light of this, there will be a major change in 2025. The complicated aggregate index will be replaced by a simple threshold: R1 institutions have more than $50 million in total research spending and award at least 70 research/scholarship doctorates (American Council on Education, 2023).

To visualize how far institutions are from the R1 cutoff, Figure 4 plots number of research doctorates against total research expenditure, with the gray dashed lines indicating the 70 doctorates and $50 million cutoffs. Note that all data in this section is based on 2019–2020 data since this is the same data used to make 2021 classification decisions. Even though R&D spending could be updated to 2021 numbers from the HERD survey, there is data selection on only the top expenditure institutions.

Under the 2025 classification, all institutions in the upper right quadrant above the two dashed lines will be R1. Current R1, R2, and MIs are delineated by different colors, with some institutions labeled. There is heavy clustering for low levels of R&D spending and awarded doctorates, which implies a long right-tail distribution for both variables. While the majority of current R2 institutions (113) will still stay as R2 under the new classification, there are 17 R2 institutions that will be classified as R1 in 2025 according to the transition matrix in Table 4. There are 8 current R1 institutions that will be downgraded to R2.

HBCUs do not have any changes in designation based on the new classification as seen in Table 5. However, Figure 5 shows institutions that are close to the new R1 thresholds, which could potentially be reclassified as R1 by the 2025 release. Howard University is very close to the

---

5 Johns Hopkins is an outlier with more than $3 billion in R&D spending, which skews the graph if included.
cutoff, with more than 100 research doctorates and more than $40 million in R&D in 2019–2020. North Carolina A&T State awarded nearly 70 research doctorates and spent more than $35 million in R&D in 2019–2020. Florida A&M University spent more than $40 million in R&D, though awarded less than 35 research doctorates. There is a cluster of three institutions (Morgan State, Jackson State, and Tennessee State Universities) that awarded well over 70 research doctorates in 2020, but spent less than $20 million in R&D.

Figure 6 zooms in on HSIs (excluding Texas A&M, which has more than $1 billion in R&D spending) and Table 6 reflects the overall transition matrix of HSIs. There are currently two R2 institutions (New Mexico State and San Diego State Universities) that would be classified as R1 institutions under the new 2025 designation based on 2019–2020 data. CUNY (City University of New York) Graduate School and University Center would be downgraded to R2.

**KEY TAKEAWAYS**

First, the number of HBCUs and HSIs entering the CCIHE system has increased from 2005 to 2021, with the number of HBCUs doubling from 8 to 16 and the number of HSIs increasing from 22 to 38. However, there are no HBCUs designated as R1 as of 2021, while HSIs have 11 R1 institutions as of 2021 (compared with 5 in 2005). It is important to keep in mind that HBCUs are historical designations, while HSIs are based on thresholds in student composition. Thus, some HSIs are large state schools with already robust research infrastructure and R1 classification.

The new 2025 designation will be clarifying for institutions that are looking to transition from R2 to R1. Whereas R2 had always been defined by thresholds in R&D spending and awarded doctorates, the confusing aggregate index for R1 designation was opaque. For example, Howard University is very close to the new R1 thresholds and will likely be reclassified as R1 in 2025. Two HSIs that are currently R2 will be reclassified as R1 in 2025, with more institutions that are close to the cutoff (e.g., Texas State University).

When considering R1 and R2 designations, it is important to step back and ask what the

---

6 Based on the 2021 HERD survey, Howard University is the only HBCU to spend more than $50 million in R&D and will thus be reclassified as R1 in 2025.
designation means for the institution. How much more funding or prestige or how many more researchers does being designated an R1 bring for a school compared with being an R2 or R3? How much weight does the Carnegie Classification carry in the future of the institution’s ability to build research capacity or secure funding? The answers to these questions are the crux of why the CCIHE would be important to consider and potentially have as a long-term objective. The answers may vary for each type of institution. For example, TCUs are not classified at all, and the objective of TCU research building will need to be defined. While the Carnegie Classification is a useful metric for understanding the current state of research, policymakers should keep in mind the role of the metric in the overall research landscape and the institutions’ specific objectives.
4

How Far are MIs from R1s and R2s?

The role of geographic spillovers may be critical in building research capacity, warranting a deeper understanding. IPEDS data provides the exact location of each institution, which can be used to calculate the distance between MIs and R1s/R2s. Ideally, there would be publicly available data on the location of all federal research facilities and the research output of each facility. Currently, there is only city data available for federal research facilities, and while research output (e.g., publications) is available from external sources such as Scopus, the data and analysis required to match institutions is far beyond the scope of this paper. Therefore, this section briefly presents preliminary findings from the available IPEDS data.

Table 7 summarizes the median distance between MIs and an R1 or R2 university. Most MIs are not close to an R1 or R2 university. Disregarding own-institution status, HBCUs are the closest with a median distance of 30 miles to an R1 and 50 miles to an R2. Note that the median distance between a TCU and an R1 is 174 miles. Around 30 percent of HBCUs have an R1 within 5 miles, while only 12 percent of HSIs have an R1 within 5 miles. There is only one TCU with an R1 located within 5 miles.

To see where MIs and R1s/R2s tend to be colocated, Figure 7 displays counties with only R1s/R2s (turquoise), only MIs (magenta), and both (dark purple). In general, MIs and R1s/R2s are not in geographically similar parts of the country. Most counties with colocation are in the West Coast (e.g., California and Arizona) with pockets of colocation in Texas, Florida, and other areas of the South.

There are a few caveats to this preliminary analysis. First, all tables and figures discount own-institution status. For example, the analysis is based on looking for the closest R1 institution to Texas A&M that is not itself. However, Texas A&M is already an R1 research institution that may not require spillovers from other R1s. Second, there are no outcomes data to empirically verify if colocation with an R1 or R2 augments research output for MIs. Do MIs and R1s collaborate when they are colocated? On the other hand, are there competitive effects for winning over faculty and students between colocated institutions? Future studies could incorporate collaborative research outputs or faculty/student mobility data to directly measure the
effects of geographical proximity between MIs and R1/R2 institutions on research capacity.

Apart from collocating with R1s/R2s, location data on federally funded R&D centers (FFRDCs) would be the next step in understanding the role of geographic spillovers in building research capacity. Currently, the publicly available data specifies FFRDCs to the city level. Together with data on outcomes (e.g., publication in certain fields) and collaboration networks, the role of geographic spillovers can be better understood in subsequent analysis. This is an important and policy-actionable question that warrants its own analysis.
R&D Expenditure

R&D EXPENDITURE DISTRIBUTIONS

R&D expenditure is extremely right-skewed for all institution types and fields. Figure 8 presents the distribution of total R&D expenditure across institution types in 2021. While R&D spending is relatively much less skewed for R1s, the right tail is still very long (and the figure is winsorized at the 99th percentile of overall spending). R1s spent much more on R&D than R2s, such that the R2 distribution is more closely aligned with HSI spending. HBCU spending is the lowest, with a much shorter right tail than HSIs, R2s, and R1s. Note that HSIs include those classified as R2 and R1, which leads to a heavy right-tail skew. Importantly, this figure is not the universe of HSIs and HBCUs, but only those that are present in the HERD survey. Thus, only the highest-spending institutions are shown. Furthermore, not all institutions in the HERD survey report spending for all fields, generating even more selection bias.

This pattern broadly holds for all fields. Figure 9 shows the kernel density for life sciences. R1s have much higher life sciences spending across the distribution than R2s, HSIs, and HBCUs. The distributional difference between R1s and other institution types is smaller for geosciences. In fact, HSIs spent more than R2s in geosciences. However, also note that many fewer institutions report any geosciences spending in the first place (except for R1s). The remaining S&E fields (engineering, physical sciences, computer science, and mathematics) in Figures 10 and 11 show the similar distributions for R2s and HSIs, with R1s spending much more across the distribution. This is also true for psychology and the social sciences in Figure 12.

It is thus difficult to summarize field-level expenditures as a single number. The sample of schools vary across fields, the sample size is not large, and most importantly, the distribution is heavily right-skewed. The median is a better measure of central tendency than the mean; therefore, for all subsequent research expenditures, this paper uses the median expenditure in figures and tables.

---

7 TCUs are excluded from the distributional figures, since there are only five TCUs in the HERD survey.
8 Other possible metrics include taking the log transformation or alternative quantiles. In the context of
R&D EXPENDITURE BY FIELD IN 2021

Total R&D Spending

Table 8 compares the median R&D spending across different fields and institution types in 2021. The median R1 spent more than $330 million in total R&D, which is over 10 times more than the median R2 that spent $25 million in total R&D. Note that the median R1 spent much more in R&D than the median R2 across all fields. The median HBCU and HSI spent $6 million and $4 million in total R&D, respectively. This is more than the median R3, which spends $1.4 million in total R&D. The median TCU spends slightly less than $600,000 in total R&D.

Non-Minority Institutions

Comparisons across fields is more intuitively depicted by bar graphs that break down total R&D spending by federal versus nonfederal sources. Across all institution types, life sciences is the most costly field. The median R1 spends more than $150 million in the life sciences, over 3 times more than the next most costly field, engineering, as illustrated in Figure 13. Physical sciences and non-S&E spending are the next most costly fields for R1s. While the median R2 also spends the most in life sciences (more than $6 million), the differential between life sciences and engineering is much smaller, as seen in Figure 14. Most fields receive R&D funds at similar levels from both federal and nonfederal sources except geosciences, which is entirely federally funded. Figure 15 pools R3 and other (non-MI) institutions together. Life sciences is the most costly field at $300,000 for the median institution, followed by engineering. However, note that the number of institutions reporting engineering expenditure is much lower than life sciences. For crossfield- and institution-type comparisons, the log transformation becomes difficult to interpret. Alternative quantiles are not used for R&D expenditure because it is not as intuitive, and there is already a selection bias for each field regarding which institutions are present (except for R1s).

9 Recall that HSIs have a longer right tail across all fields; there is large heterogeneity in the sizes and types of HSIs.

10 Nonfederal funding is broken down into state/local funding for only the long HERD survey, which would exclude a number of smaller institutions filling out the short HERD survey.

11 The tail end of spending is characterized by the top 10 R1s by R&D expenditure in Table 20. There is some differential spending across fields as seen in Table 21. For example, Johns Hopkins is in the top 3 for R&D spending in life sciences, computer science, engineering, mathematics, and physical sciences, but does not enter the top 10 for psychology or social sciences.
all three institution groups, mathematics and statistics has the lowest levels of R&D spending.

Historically Black Colleges and Universities

Figure 16 shows a similar trend of R&D spending for the median HBCU. Life sciences dominate at more than $3 million in spending, followed by the physical sciences and engineering at less than $1 million each in R&D. Overall, there is a higher fraction of federally funded R&D spending across fields for the median HBCU. Note the spending differential at the tail end in Figure 17, which breaks down the R&D spending of the top 10 HBCUs. The top three institutions (Howard, Florida A&M, North Carolina A&T State) spent more than $40 million each in R&D for 2021. The top five HBCUs spent heavily in the life sciences, with substantial spending in engineering as well. Florida A&M and North Carolina A&T State also spent heavily in non-S&E R&D. Florida A&M, Morgan State, and the University of the Virgin Islands also spent substantially in geosciences in 2021, and Tuskegee University spent substantially in the social sciences. Table 23 ranks the top 10 HBCUs by R&D spending in each field. There is a certain degree of specialization within fields. For example, Howard spent the most in total R&D, but does not rank in the top 5 spending for geosciences or mathematics and statistics.

Hispanic-Serving Institutions

The spending distribution across fields is quite different for HSIs. Figure 18 shows that the median HSI spent the most in engineering at around $1.5 million, followed by non-S&E spending, then the life sciences. There is also substantial spending in the geosciences compared with other fields, which is also entirely federally funded. HSIs are also characterized by a very long right tail as seen in Figure 19, where the top 10 HSIs each spent more than $100 million in total R&D. University of New Mexico and UC Riverside spent the majority of their R&D in the life sciences, while the other top HSIs spent the most R&D engineering and physical

---

12 Refer to Table 22 for exact quantities in spending for the top 10 HBCUs.
13 However, there are only 46 HSIs reporting engineering R&D, while there are 72 HSIs reporting life sciences R&D.
14 Texas A&M spent more than $1 billion in total R&D, so it is exclude, as it skews the graph. Refer to Table 24 for the R&D spending breakdown of Texas A&M and the other top nine HSIs across fields.
There is more homogeneity in HSIs that spent the most across all fields compared with HBCUs. Table 25 shows permutations of Texas A&M, University of New Mexico, UC Santa Barbara, and Florida International University as the top spending HSIs in each field.

**Tribal Colleges and Universities**

Figure 20 seems to show that TCUs are heavily skewed toward spending in the social sciences; however, this is due to the small sample size for each field. Most fields only have one or two TCUs reporting R&D spending, except for geosciences and life sciences, which have three and five TCUs, respectively. Nearly all reported R&D spending is from federal grants. Figure 21 shows the R&D breakdown for all five TCUs in the HERD survey. Northwest Indian College spent nearly $3 million in total R&D in 2021, with the majority of spending in geosciences and social sciences. On the other hand, Diné College spent the entirety of its reported $1.5 million R&D on life sciences. The dearth of information for MIs, especially for TCUs, is a major limitation in understanding the current state of research capacity.

**Total R&D Spending per Student**

One important consideration is how much R&D is being spent per student. For large schools, higher R&D spending may be mechanical (e.g., more graduate student stipends). Tables 9 and 10 show the median R&D expenditure per undergraduate and graduate student, respectively. Student enrollment statistics are based on fall 2020 data, so there is no differential weighting across fields within an institution.

Though R1s still spent the most per student, the differential is smaller with HBCUs and R2s after re-weighting. The median R1 spent $18 million on each undergraduate student, while the median HBCU and R2 spent more than $2 million per undergraduate student. The median TCU spent more than $1 million per undergraduate student. However, the median HSI spent only $650,000 per undergraduate student, and the median R3 spent even less per undergraduate student at $320,000. Adjusting for undergraduate student population, HBCUs and TCUs spent

---

15 The correlation between undergraduate and graduate enrollment across all, non–Special Focus Institutions is 0.54; hence, both weighting schemes are presented.
relatively more per student (on par with the median R2) than HSIs and R3s. On average, HSIs have much larger student populations than HBCUs and TCUs. The distribution of R&D spending across fields within institution types remains relatively constant as the unweighted case.

Institutions have much fewer graduate students compared with undergraduates, and the spending gap is even smaller between R1s and other institution types after adjusting for the number of graduate students. The median HBCU and TCU spent $16 million per graduate student, compared with the median R1, which spent $54 million per graduate student. This is much higher than the median HSI and R2, which spent $4 million and $9 million per graduate student, respectively. Spending in the geosciences between the median HBCU, HSI, and R1 are approximately the same at $400,000 to $500,000. Median spending per student in HBCUs for mathematics and physical sciences are approximately half that of R1s, which is a much smaller differential in the alternative weighting cases. However, it is important to consider that while nearly all R1s are in the HERD survey, there is immense selection on other institution types to be the highest spending institutions. Weighting by graduate students is only possible for institutions with graduate programs, so Table 10 has an additional layer of selection bias.

**CAPITALIZED EQUIPMENT EXPENDITURE BY FIELD IN 2021**

Institutions in the HERD survey also reported a breakdown of their R&D spending, one of which is expenditure on capitalized equipment. Distributional data for equipment spending is not shown, though it is similarly right-skewed as total R&D spending. Thus, the median is reported for all subsequent analysis.

Table 11 reports the median capitalized equipment expenditure across different fields and institution types in 2021. The median R1 spent slightly less than $10 million on capitalized equipment, much more than the median R2, which spent more than $800,000. The median HBCU, HSI, and TCU all spent slightly more than $200,000. The median institution did not spend any money on capitalized equipment across most fields except for R1s. The median R1 spent more than $1 million in capitalized equipment for life sciences, engineering, and physical sciences. Fields such as mathematics, psychology, and social sciences generated nearly no
capitalized equipment expenditure even for R1s.

Tables 12 and 13 display the median capitalized equipment expenditure per undergraduate and graduate student, respectively. The differential between the median R1 and MI is smaller adjusting for the number of students. Per undergraduate, the median TCU spent nearly $400,000 on capitalized equipment, which is on par with the median R1 spending per undergraduate at more than $500,000. Note that for the median TCU, this substantial spending was from geosciences and life sciences.16 The median HBCU and HSI spent more than $500,000 and $150,000 on capitalized equipment, respectively, per graduate student. This is on par with the median R2 spending of more than $250,000 per graduate student. R1s spent slightly more on capitalized equipment per graduate student at nearly $1.5 million. Across all institution types, the most costly fields in terms of equipment per graduate student are life sciences, engineering, and physical sciences.

KEY TAKEAWAYS

This section summarizes the major takeaways from the R&D expenditure analysis:

1. **Skewness of Expenditure Data**: R&D expenditure is extremely right skewed for all cuts of institution type and field. Within each group of field and institution type, a small number of institutions spent the majority of R&D. Thus, all analysis uses the median as opposed to the average. Even with the median, it is important to keep in mind that there are outliers (e.g., R1s and HSIs tend to have longer right tails).

2. **Spending Distribution Across Institutions**: The median R1 spent magnitudes more on R&D compared with any other institution type, with the median R2 being much closer to the median HSI than the median R1. Due to the limited data on TCUs (only five institutions in total, and only life sciences expenditure was reported by all five), it is more difficult to make representative comparisons.

3. **Field-Specific Expenditure Patterns**: Life sciences dominate R&D spending across most institution types, followed by notable spending in engineering and physical sciences. The median HBCU spent relatively more on the life sciences compared

---

16 Since there are only five TCUs in the HERD survey, this spending comes from Northwest Indian College.
with the median HSI, which spent relatively more on engineering. Mathematics had the lowest levels of R&D spending across all institution types. This is followed by geosciences, computer science, and psychology. Note that geosciences is almost entirely federally funded.

4. **Per-Student Spending**: While the median R1 still spent the most per student, the disparity with the median HBCU and HSI is much smaller after adjusting for number of students. The median HBCU and TCU spent relatively more per student compared with the median HSI and R3. The spending disparity is even narrower when adjusting for the number of graduate students (which uses a subsample of the data, since not all institutions have graduate students).

5. **Equipment Expenditure**: Capitalized equipment spending is also very right-skewed, with the spending being zero for the median institution across most fields and institution types (except R1s). Life sciences, engineering, and physical sciences are the most costly fields in terms of capitalized equipment expenditure overall and per student. Mathematics, psychology, and social sciences generated almost no capitalized equipment expenditure for all institution types.
Facility Space

FACILITY SPACE DISTRIBUTIONS

Unlike reports of R&D spending in the HERD survey, the S&E Research Facilities survey maintains a balanced panel of institutions across fields. Thus, the comparison across fields is more interpretable, though the sample size is much smaller. Figures 22–26 show the kernel densities of each field for R1s, R2s, HSIs, and HBCUs. Facility space is also very right-skewed for all institution types and fields. R1s tend to have very long right tails and distributions shifted toward higher levels of facility space than R2s, HSIs, and HBCUs. HSIs and R2s have similar distributions of facility space, though there are only 25 HSIs in the S&E Research Facilities survey, so there is severe selection bias. HBCUs have much lower levels of facility space across fields compared with R1s, R2s, and HSIs.

The difference in facility space between R1s and other institution types is especially large for the biological and health sciences, as seen in Figure 22. Figure 23 shows a smaller distributional difference for natural resources and geosciences. In fact, the 25 HSIs have a similar distribution of geoscience facility space as the 76 R1s in the S&E Research Facilities survey. Figure 24 also shows a long right tail for R1s in engineering and physical sciences, though HSIs have a substantial amount of physical sciences facility space as well. HBCUs and R2s have a similar distribution for computer science facility space, followed by HSIs then R1s with the most space, as seen in Figure 25. The difference between R1s and other institution types is also smaller for mathematics facility space, where HBCUs have the longest right tail. R2s and HSIs have similar amounts of facility space for psychology and social science, as seen in Figure 26, which is slightly less than the facility space distribution of R1s. However, HBCUs have significantly less facility space for psychology and social sciences with a large density near zero.

Similar to R&D expenditure, facility space is highly skewed and the sample size is even

---

17 TCU's are excluded from figures of facility space because there is only one TCU in the Facilities survey.
18 There are also only 81 out of 110 R2s present in the Facilities survey. Recall that Special Focus Institutions are excluded from analysis.
19 This excludes space for medical schools and clinical trials.
smaller. The panel is balanced across fields; however, this also leads to a high number of zeros because many institutions do not offer all fields. It is not rare for even the median to be zero for certain subgroups regarding facility space. For all subsequent analysis, this paper shows both the median and the 75th percentile of facility space for each institution type and field.

FACILITY SPACE BY FIELD IN 2021

Total Facility Space

Table 14 compares the median facility space (in square feet) across different fields and institution types in 2021. Since facility space is quite specialized across fields, a pooled total is not reported. There is substantial heterogeneity across fields, with many fields reporting zero square feet for the median institution type. For example, the median institution for all types except HBCUs and TCUs reported zero square feet for agricultural sciences (including R1s). Across all other fields, the median R1 has significantly more facility space. Biological sciences and health sciences occupy the most space, followed by engineering and physical sciences. Geosciences is the next highest in terms of facility space for R1s. The following sections examine cross-field comparisons in more detail graphically.

Non-Minority Institutions

Figures 27–29 show the median and 75th percentile of facility space for R1s, R2s, and other (includes R3) institutions across fields. All figures depict current and planned facility space; however, there is no planned facility space for any field across all institution types at the 75th percentile. Biological sciences require the most facility space across institution types and at all percentiles, followed by physical sciences across the board. For sense of scale, the median R1 has nearly 250,000 square feet devoted to biological sciences, with the 75th percentile at more

---

20 A large part is due to the severe sample selection issue in the HERD survey.
21 Planned construction is non-zero for only a few institutions in the far right tail. Unlike R&D expenditure, facility space is not as adjustable year to year, and will require large, discrete investments that take time.
than 400,000 square feet. R2s spent much less—the 75th percentile R2 has less than 80,000 square feet for the biological sciences. Engineering took up the most space following biological sciences for R2s as well as R1s (discounting health sciences, which is also significant for R1s). Engineering is relatively space intensive for non-R1s and R2s only at the 75th percentile. In general, the relative ranking of fields by facility space is consistent for the median and 75th percentile institutions within each type. The median R1 and R2 has relatively much less facility space for mathematics and agricultural sciences and natural resources. Geosciences also occupy much less space, on a similar order to that of psychology and social sciences.

Historically Black Colleges and Universities

Figure 30 shows the current and planned facility space for HBCUs at the median and 75th percentile. There are no planned facility spaces at the 75th percentile, similar to non-MIs in the S&E Research Facilities survey. The median HBCU in the sample of 38 institutions present in the Facilities survey shows the biological sciences take up the most space at around 10,000 square feet, followed by physical sciences at 5,000 square feet and engineering, computer science, and agricultural sciences (at around 2,000 square feet). While the relative ordering of fields is consistent in the 75th percentile case, there is a surge in agricultural sciences space to 40,000 square feet, overtaking biological sciences. There are also increases in health sciences at the 75th percentile. Figure 30 reveals substantial heterogeneity in how facility space is used across fields. Howard has 300,000 square feet for biological and health sciences, while Fort Valley State and Prairie View A&M Universities have nearly 200,000 square feet for just agricultural sciences. Tuskegee has a substantial amount of space for both engineering and agricultural sciences, while Delaware State has the highest square footage for physical sciences. North Carolina Central University has more than 200,000 square feet entirely in other sciences. Dillard University has the largest reported square footage overall distributed across many fields, with

---

22 The tail end of facility space is characterized by the top 10 R1s in terms of total facility space in Table 26. However, there are significant differences in the set of top 10 R1s when ranking within field, as seen in Table 27. For example, University of Florida has the most facility space for agricultural sciences but does not make the top 10 for biological or physical sciences. Even among R1s, there is a degree of specialization by institutions.

23 Refer to Table 28 for exact square footage for the top 10 HBCUs.
150,000 in geosciences. The extent of specialization within fields is also reflected in Table 29, which ranks the top 10 HBCUs by facility space within each field. In general, mathematics, computer science, and the social sciences take up relatively much less facility space even in the right tail of HBCUs.

**Hispanic-Serving Institutions**

There is much more of a difference in the 75th percentile versus the median distributions of HSI facility space, as seen in Figure 32. However, there are only 25 HSIs in the S&E Research Facilities survey (less than 10 percent of HSIs), so the sample is likely not representative, and there would be higher variance with the small sample. The median HSI in the survey has the most facility space for biological sciences at nearly 20,000 square feet, followed by physical sciences at 15,000 square feet. All other fields have less than 5,000 square feet and have a fairly uniform distribution of facility space (except mathematics with near zero, and zero for agricultural sciences and natural resources). However, while the 75th percentile HSI has proportionally the most spending in biological sciences (more than 50,000 square feet) followed by physical sciences (nearly 40,000 square feet), facility space is then followed by engineering (nearly 40,000 square feet), then geosciences and health sciences (nearly 25,000 square feet). The remaining fields have around 10,000 square feet or less. Mathematics, agricultural sciences, and natural resources still occupy the least facility space.

Despite the small sample, the long right tail of HSI research capacity is reflected in Figure 32, which breaks down the facility space usage of the top 10 HSIs. Three of the top four HSIs in terms of facility space belong to the UC system, with the top institution as UC Riverside at more than 1 million square feet of facility space. However, only the top six schools have more than 500,000 square feet of facility space. There is more even distribution across different fields for all the top 10 institutions, with relatively more spending in biological sciences, engineering, and physical sciences across the board. UC Riverside also has substantial agricultural science facility space, and Florida International has a substantial amount of psychology facility space. Table 31 ranks the

---

24 For unclear reasons, not all R1 HSIs are in the Facilities survey. For example, even though Texas A&M (College Station) spent the most in terms of R&D for HSIs, it is absent from the Facilities survey.

25 Refer to Table 30 for exact square footage for the top 10 HSIs.
top 10 HSIs by facility space within each field. There is less variance in the set of top five institutions that are common across fields.

**Tribal Colleges and Universities**

Figure 33 breaks down the facility space by field of Northwest Indian College, the only TCU in the S&E Research Facilities survey. Of a total of 7,400 square feet, nearly 30 percent is devoted each to biological sciences and geosciences. Agricultural sciences occupy 16 percent of facility space, and the remaining fields occupy less than 10 percent of facility space. The lack of data on TCUs makes it difficult to analyze the state of the current research capacity.

**Total Facility Space per Student**

To adjust for the number of students at an institution, Tables 15 and 16 show the median facility space per undergraduate and graduate student, respectively. The numbers are rounded to the nearest square feet. Thus, when the median institution for the majority of fields assigns zero square feet per undergraduate student, it should be interpreted as less than 0.5 square feet per undergraduate. The exception is in biological sciences, which assigns 2 to 4 square feet per undergraduate student, and up to 14 square feet per undergraduate for the median R1. When adjusting for the facility space per graduate student, the median HBCU and R1 assigns 32 to 33 square feet per student in the biological sciences. Physical sciences space per graduate student is also above 10 square feet for HBCUs and R1s. However, given that aggregate student numbers are being used and are not adjusted for students per field, “facility space per student” should be interpreted with caution.

**FACILITY COST BY FIELD IN 2021**

In addition to space, the S&E Research Facilities survey also provides information on the costs of repair and renovation and new construction. The costs of R&R and new construction are
also very right-skewed, with even more sparsity than facility space, thus both the median and 75th percentile are reported. Figure 35 shows the current and planned R&R costs for the median and 75th percentile R1s by field. The median R1 R&R costs are close to $1.5 million for the biological sciences (and even higher planned R&R costs), with costs above $500,000 for engineering, health sciences, and physical sciences. R&R costs for all other fields is zero. The 75th percentile R1 spent more than $6 million in current and planned R&R costs for biological sciences, followed by similar spending for engineering and health sciences (both current and planned). The 75th percentile R1 spent more than $2 million in current and planned R&R costs for physical sciences, and nearly zero for all other fields. For all other institutions types, there is zero current and planned R&R costs across all fields, so they are not shown.  

Figure 36 breaks down the funding source of total R&R costs (across all fields) for the median and 75th percentile institution. The median R1 and R2 fund all R&R costs via nongovernment sources, with state/local funding becoming involved at the 75th percentile. On the other hand, the median HSI funded approximately 30 percent of their R&R costs from state/local sources and the median R3 funded all R&R costs from state/local sources. This is still true for the 75th percentile HSI and R3 institution, with nearly 50 percent of R&R costs funded by state/local sources for the 75th percentile HSI. The 75th percentile HBCU is the only institution to be mostly funded by the federal government for R&R costs. Note that the median HSI and R3 in the sample have high levels of total R&R costs (more than $6 million and $3 million, respectively) but zero for each individual field. This implies the highest spending HSIs in the total amount also tend to be the HSIs spending the most in a given field (above the 75th percentile), so the distribution is extremely right-skewed.

Figure 37 shows a similar funding pattern for new construction costs. The main difference is HSIs are mainly funded by state and local sources, with a small portion coming from nongovernment funds. New construction costs are primarily driven by the tail behavior of a few institutions across all categories. The median and 75th percentile of new construction costs is zero for all institution types across individual fields, but it is non-zero when aggregated to the institution level. There are high levels of total new construction costs for the median institutions: more than $15 million for R1s, nearly $15 million for HSIs and R3s, and approaching $5 million.

---

[26] Only the 75th percentile R2 has non-zero R&R costs for the biological sciences.  
[27] Note the small sample size for both groups at only 25 and 26 institutions, respectively.
for HBCUs, R2s, and other institutions. This suggests that the institutions with highest net construction costs are also the institutions with the highest costs in a given field.

**KEY TAKEAWAYS**

This section summarizes the major takeaways from the facility space and cost analysis:

1. **Distribution of Facility Space Across Institutions:** The S&E Research Facilities survey presents a balanced panel across fields, offering more interpretable comparisons than the HERD survey. However, it has a much smaller sample size. Facility space is highly right-skewed across all institution types and fields, with R1 institutions exhibiting significantly larger facility spaces compared with R2s, HSIs, and HBCUs (such as in the biological sciences). HSIs and R2s have similar distributions of facility space, but the HSI sample exhibits more selection bias with only 25 institutions.

2. **Facility Space by Field in 2021:** There is more variation on the extensive margin for facility space and costs. For institution types across different fields, there are cases of the median institution having zero square feet of facility space. There is no planned facility space for all fields and institution types even at the 75th percentile. Across institution types, biological and physical sciences relatively occupy the most facility space for the median institution.
   - Engineering is also space intensive for R1s and R2s. Math and agricultural sciences/natural resources take up much less space for the median R1 and R2.
   - Agricultural sciences take up the most space for the 75th percentile HBCU. There is substantial specialization in fields across the top 10 HBCUs.
   - The 75th percentile HSI also has substantial facility space for engineering, geosciences, and health sciences. The sample size is quite small with only 25 HSIs, but the tail spending is very right-skewed for the top HSIs. There is more even distribution across different fields.
   - Northwest Indian College is the only TCU in the S&E Research Facilities survey, which has 30 percent of facility space devoted each to biological sciences.
3. **Facility Space per Student:** Most fields allocate less than 0.5 square feet per student (both graduate and undergraduate) except biological sciences. However, this metric should be interpreted with caution, since the number of students is not adjusted by field, which would make significant differences in facility space.

4. **Facility Costs:** Repair and renovation costs are highly right-skewed and are more sparse than facility space. R1 institutions spend significant R&R costs for biological sciences, engineering, and health sciences. R&R costs for other median and 75th percentile institutions are zero.

5. **Funding Sources for Facility Costs:** R1s and R2s mostly rely on nongovernment sources, HSIs and R3s rely partially and fully on state and local funding (respectively), and HBCUs rely mostly on federal funding.
R&D Expenditure versus Facility Space

Given the data on both R&D expenditure and facility space, the next step is to see how these two measures of research capacity are related. Institutions are matched via their IPEDS ID and field across the HERD and S&E Research Facilities surveys, so the analysis is done at the institution by field level. Table 2 shows the number of institutions present in both surveys (i.e., matched) by field. Life sciences and physical sciences have more than 300 institutions in both surveys, and all remaining fields have more than 200 institutions in both surveys (except other sciences that this paper does not focus on).

Figures 38–47 are scatterplots of R&D expenditure versus facility space for each field, with the points colored by institution type. Since the data is highly right-skewed, both Pearson and Spearman (i.e., rank) correlations are reported. Spearman correlations are less influenced by outliers, since it compares the rank of two variables. The Pearson correlation line is shown for each plot (i.e., line of best fit with an intercept of zero).

- **Computer Science:** Figure 38 shows the scatterplot for all matched institutions in computer science, though the presence of three outliers makes most of the graph uninterpretable and drives the Pearson correlation. Figure 39 zooms in on computer science facility space and R&D spending without the three outliers. The correlation is around 0.6 for both correlations, which is a moderately strong association. There are a few HSIs that spent adequately on computer science R&D and facility space, and most HBCUs spent relatively less on computer science R&D despite having more facility space.

---

28 Reported space for biological sciences, health sciences, medical schools, agricultural sciences, and natural resources are aggregated to correspond to life sciences in the HERD survey. Clinical trial space is excluded.

29 Tables 32 and 33 list all HBCUs and HSIs present in the HERD and Facilities surveys. Note that there are zeros in the raw data, which is different from missing, and they are retained.
• **Engineering:** Though the data is right-skewed as seen in Figure 40, there are no major outliers that drive the result. The Pearson and Spearman correlations are 0.81 and 0.88, respectively, which reflects a very strong association between engineering facility space and R&D spending. There are five HSIs that spent substantially in engineering R&D and facility space in proportion to the Pearson correlation. HBCUs are clustered near zero, with two HBCUs that have relatively more facility space but proportionally less R&D spending.

• **Geosciences:** Geosciences also have a fairly strong association between facility space and R&D spending, as seen in Figure 41. The Pearson and Spearman correlations are 0.77 and 0.74, respectively. There are three HSIs and one HBCU that have substantially more facility space than the remaining MIs, though they all spent proportionally equal to or less in R&D than suggested by the correlation.

• **Life Sciences:** Figure 42 shows a very strong association between life sciences facility space and R&D spending, with Pearson and Spearman correlations of 0.85 and 0.82, respectively. Note the immense right tail of life sciences for both facility space and R&D spending, with the majority of MIs clustered near zero. There is one HSI that has substantially more facility space than the remaining MIs.

• **Mathematics and Statistics:** The association between facility space and R&D spending is much weaker for mathematics and statistics, as seen in Figure 43. The Pearson and Spearman correlations are 0.45 and 0.35, respectively. Notably, there are many institutions with less facility space but substantial R&D spending, as is shown by the bulk lying above the Pearson correlation line. There are a few HSIs with more facility space and less R&D spending, and one HBCU with the second-highest facility space but nearly no R&D spending.

• **Physical Sciences:** Figure 44 shows a very strong association between physical sciences facility space and R&D spending, with Pearson and Spearman correlations of 0.90 and 0.76, respectively. The large difference between the Pearson and Spearman correlation is driven by outliers, which can be seen with the highest-spending institution having relatively less facility space. Since the tail is extremely long, Figure 45 zooms in on institutions with less than 200,000 square feet of facility space (the top facility space is approximately 500,000 square feet). This encompasses all MIs, with
three HSIs exceeding 150,000 square feet of space. Once again, while HSIs are proportionally distributed above and below the Pearson correlation line, HBCUs with more facility space tend to spend proportionally less in R&D.

- **Psychology:** Psychology exhibits a moderately strong correlation between facility space and R&D spending, as seen in Figure 46. The Pearson and Spearman correlations are 0.62 and 0.64, respectively. Similar to mathematics and statistics, there is a bulk of institutions with less than 50,000 square feet that spend substantially in R&D. Note that there are a few HSIs in the far right tail of both facility space and R&D spending, whereas most HBCUs are clustered near zero.

- **Social Sciences:** Similar to psychology, social sciences exhibit a moderately strong association, with Pearson and Spearman correlations of 0.62 and 0.53, as seen in Figure 47. The highest spending institution is an outlier in having proportionally less facility space. The remaining institutions fall near the Pearson correlation line. Apart from one HSI and one HBCU, most MIs that have more facility space spent proportionally less in R&D, especially for HSIs.
• **Other Sciences:** Figure 48 has only 60 matched institutions for other sciences, which is a small sample. The correlation is weak, with Pearson and Spearman correlations of 0.38 and 0.41, respectively.

**QUANTILE REGRESSION OF R&D EXPENDITURE ON FACILITY SPACE**

Due to the skewed nature of spending and facility space data, this paper uses quantile regressions to estimate the relationship between R&D expenditure and facility space, adjusting for graduate student counts in Table 17. The coefficients are purely correlational and not causal. The 25th, 50th, and 75th quantiles of R&D expenditure are estimated (in 000’s of $) on facility space and total graduate enrollment for each field.  

The interpretation of the quantile regression coefficients is the same as that of the OLS regression coefficients, but the coefficients are estimated at different points of the conditional distribution of R&D expenditure. For example, the 25th quantile regression coefficient is the effect of a one-unit increase in the independent variable on the 25th percentile of R&D expenditure (in 000’s of $).  

Note that the independent variables are also skewed, so bootstrap standard errors with 100 replications are used to take into account the heteroskedasticity.

Across all fields, one additional square foot of facility space is associated with a larger increase in R&D expenditure at higher percentiles. This suggests increasing economies of scale for institutions that already have high research capacity. The pseudo $R^2$ also increases at larger percentiles across all fields. This is not surprising given that the right tails of both the space and expenditure distributions have much less variance.

Computer science, engineering, physical sciences, and life sciences have the largest coefficients for facility space across all quantiles. At the 25th percentile of R&D expenditure, an additional square foot of facility space is associated with at least a $150 increase in R&D

---

30 Unfortunately, graduate student enrollment is pooled to the institution level, so it cannot be adjusted for the number of students by field. The regression is run with all institutions, since there would not be enough power to run by institution type to examine heterogeneous effects.

31 Facility space is measured in square feet and graduate enrollment is measured as total number of enrolled graduate students. Graduate student enrollment is used to control for institution size; however, interpretation is difficult, since it is not graduate students per field. The discussion is thus focused on the relationship between space and expenditure.
spending for these four fields (except life sciences, which is a $130 increase). At the 50th percentile, an extra square foot is associated with approximately a $250 increase in R&D (though computer science stands out at above $350). Facility space has the largest effect on computer science R&D expenditure, with one extra square foot suggesting a $359 and $629 increase in R&D spending at the 50th and 75th percentiles, respectively. Life sciences is the second highest at the 75th percentile, with an additional square foot suggesting a $411 increase in R&D spending. Engineering and physical sciences follow suit at the 75th percentile of R&D expenditure, with an additional square foot suggesting a $349 and $343 increase in R&D spending, respectively. The pseudo $R^2$ is consistently the highest for life sciences, engineering, and physical sciences across all quantiles. Then controlling for graduate student size, the facility space variation explains around half (or more) of the variance in median and 75th percentile R&D expenditure for these three fields.

The remaining fields (i.e., geosciences, mathematics, other sciences, psychology, social sciences) have much lower associations between facility space and R&D expenditure at the left tail of the distribution. At the 25th percentile, an extra square foot in space is associated with less than $100 in R&D spending. For mathematics and other sciences, the association is not statistically significant. At the 50th percentile, an extra square foot is associated with a $130 increase in R&D spending for geosciences and psychology, and even higher at $173 for social sciences. Mathematics remains the lowest (aside from other sciences) at $103. At the 75th percentile, an extra square foot is associated with a $290 increase in R&D spending for geosciences and social sciences, with psychology lagging at $204 and mathematics in last place again at $150 (aside from other sciences). Facility space variation explains relatively less of the variance in R&D expenditure for math, with the highest pseudo $R^2$ at the 75th percentile being 0.36. For all other fields (aside from other sciences), the variation in facility space explains at least 40 to 50 percent of the variance in R&D expenditure at the 75th percentile.

---

32 Recall that other sciences have a very low match rate of only 60 institutions, which limits the power of the regression.
KEY TAKEAWAYS

This section summarizes the major takeaways from the R&D expenditure versus facility space analysis:

1. **Space and Spending Correlation by Field:** There is a very strong association between facility space and R&D expenditure for engineering, geosciences, life sciences, and physical sciences. There is a moderately strong association for computer science, psychology, and social sciences. There is a weak association for mathematics and statistics.33

2. **Space and Spending Correlation for MIs:** HBCUs are clustered closer to zero across most fields. For fields where there are HBCUs with relatively more facility space, there is also proportionally less R&D spending. HSIs are more evenly distributed across the Pearson correlation line, with a few HSIs that spend substantially in R&D and facility space. There are a few HSIs in the right tail of facility space and R&D spending for engineering, physical sciences, geosciences, and psychology.

3. **Space and Spending Controlling for Graduate Enrollment:** Based on quantile regressions with bootstrapped standard errors to account for distribution skewness, the association between facility space and R&D expenditure is stronger at higher percentiles of R&D expenditure. This suggests economies of scale for institutions with higher research capacity.

4. **Space and Spending Controlling for Graduate Enrollment by Field:** The effect of facility space on R&D expenditure is largest for computer science (especially), life sciences, engineering, and physical sciences. This is not surprising for the latter three fields given the distributions of equipment expenditure and facility space. The effect is the lowest by far for mathematics (aside from other sciences), followed by psychology, social sciences, and geosciences.

33 This is also the case for other sciences, but the matched sample size is too small to draw conclusions from.
R&D Expenditure and Facility Space Over Time

It is possible to examine the trends of R&D expenditure and facility space over the last decade. The HERD survey has annual data from 2012 to 2021, and the S&E Research Facilities survey has biennial data from 2011 to 2021. One major reason for change over time is due to an unbalanced panel of institutions that enter and exit the surveys. Tables 18 and 19 show the percent of institutions that are present in the previous year’s data for the HERD and S&E Research Facilities surveys, respectively. The panels are well balanced across institution types, especially in the HERD survey and 2015 onward in the S&E Research Facilities survey. Given that the institution sample remains quite similar over time, this paper focuses less on the time trends.

R&D EXPENDITURE FROM 2012 TO 2021

Figures 48–58 show the median R&D expenditure for each field and split by institution type. R1s spend much more than all other institution types in all fields, which distorts the axes of the graphs. Thus, the left-hand side graph compares the median expenditure for all institution types except R1s, and the right-hand side graph compares R1s, R2s, and other non-MIs. Both the red and orange lines in the left-hand and right-hand side graphs depict median R2 expenditure, respectively. The question of what consistent R&D expenditure implies for the institution remains. How much is attributed to recurring costs (i.e., human labor) versus fixed costs (i.e., capital), which may persistently increase research productivity? It may also be the case that increasing research productivity due to capital expenditure also raises levels of recurring costs if more researchers then come to the institution. The current analysis cannot answer this question without microdata on research expenditure and personnel.

Figure 48 shows how much more the median R1 spends in total R&D expenditure compared with even the median R2, which spends significantly more than the median HSI, HBCU, TCU, and other non-MIs. Total R&D spending has remained relatively constant over time.

---

34 “Other” includes R3s.
for all institution types except R1s, which has increased from $250 million to more than $300 million.

- **Computer Science**: Figure 49 shows that both the median R1 and R2 have increased computer science R&D expenditure in the last decade. The median R1 has increased spending from less than $6 million to more than $8 million and the median R2 has increased spending from less than $600,000 to nearly $1 million. The median HSI and R2 spent similar amounts up until 2015, after which median HSI spending slightly decreased in computer science. The median HBCU expenditure remained relatively constant at around $200,000.

- **Engineering**: Figure 50 shows that while the median R1 has increased engineering R&D expenditure from $40 million to $50 million in the last decade, the median R2 spending has remained relatively constant at around $5 million. The median HSI and HBCU have slightly increased over time from $2 million in 2012.

- **Geosciences**: Expenditure over time is unique, since the median R1 has decreased spending from more than $5 million to less than $4 million, as seen in Figure 51. The median TCU had a spike in spending to $800,000 in 2019, but currently spend slightly more than $200,000. The median HSI has been volatile in terms of geosciences R&D spending, and currently spends more than $600,000, which is higher than the median R2. The median HBCU decreased spending over time to around $200,000 in 2021.

- **Life Sciences**: R1 expenditure in the life sciences is the most out of any field, increasing from $125 million to more than $150 million, as seen in Figure 52. The remaining institution types spend much less, with the median R2 increasing spending from less than $4 million to more than $6 million. The median HBCU then spends the most at more than $3 million, followed by the median HSI at around $1 million.

- **Mathematics and Statistics**: Figure 53 shows much lower levels of spending in mathematics across the board, with R1s increasing from $2 million to $2.5 million. The median R2, HSI, and HBCU actually spent similar amounts in 2012 at $200,000 to $250,000. However, median R2 spending remained constant while the median HBCU and HSI decreased mathematics R&D expenditure to less than $150,000.

- **Physical Sciences**: Figure 54 shows the median R1 spending significantly more in physical sciences, increasing from $15 million to $20 million in the last decade. The
median R2 increased slightly from $1.5 million to $2 million. The median HBCU and HSI spent significantly less at less than $1 million. The median TCU had a spike in spending in 2018 at $1 million, then decreased rapidly.

- **Psychology:** Figure 55 shows that the median R1 has increased psychology R&D expenditure from $4 million to nearly $6 million in the last decade. The median R2 has remained relatively constant at around $500,000, and the median HSI has oscillated between $200,000 and $400,000 in R&D. The median HBCU spent less than $100,000 until 2020, when it increased to more than $200,000.

- **Social Sciences:** The median R1 has increased social sciences R&D expenditure from more than $5 million to more than $10 million, as seen in Figure 56. The median R2 has slightly increased spending to less than $1 million, while the median HSI and HBCU spent less than $500,000. The median TCU experienced a spike in spending in 2013 at $2 million, and in the last 2 years at more than $1 million.

- **Other Sciences:** Figure 57 also shows decreased spending from the median R1 in other sciences, from $4 million to less than $3 million. The median HSI spending peaked in 2012 at more than $1 million, but has remained relatively constant around $400,000. The median R2 and HBCU spent similar magnitudes in other sciences.

- **Non-Science and Engineering:** Non-S&E covers a wide breadth of fields (e.g., humanities, arts). The median R1 has increased R&D expenditure significantly in non-S&E fields, from $10 million to more than $20 million, as seen in Figure 58. The median R2 has also increased spending in non-S&E fields from $2 million to $3 million, while the median HSI and HBCU remained relatively constant at $1 million and $500,000, respectively.

**Capitalized Equipment Expenditure Over Time**

Capitalized equipment expenditure trends follow that of total R&D expenditure with much lower dollar amounts, and thus higher volatility. Specific trends to note in those figures are as follows.

Despite increasing expenditure for the median R1, computer science capitalized
equipment does not cost too much in dollar amounts according to Figure 72 (hundreds of thousands of dollars). Mathematics and statistics have no capitalized equipment expenditure in Figure 76. This is in contrast to engineering and life sciences in Figures 73 and 75, where the median R1 spends $2 to $3 million just on equipment. Physical sciences is in a range of $1 million for median R1 equipment expenditure in Figure 77, which is better but still significantly more than the median R2 or MI.

TCUs exhibit a steady and sharp increase in equipment expenditure for geosciences in Figure 74, which is around half of the median R1 expenditure at more than $150,000. A similarly sharp increase in spending is true for TCUs in social sciences and non-S&E, as shown in Figures 79 and 81.

For psychology, social sciences, and non-S&E, median R1 equipment spending over time is on the order of $20,000 to $40,000, which is much less than that of science, technology, engineering, and mathematics fields (see Figures 78, 79, and 81).

**FACILITY SPACE AND COSTS FROM 2011 TO 2021**

Since there is only one TCU in the S&E Research Facilities survey, this paper compares the time trends in facility space directly for HSIs, HBCUs, R1s, and R2s in each field, shown in Figures 59–69. Six cycles of the S&E Research Facilities survey from 2011 to 2021 is used. Due to both the skewness and sparsity of that survey, the left-hand and right-hand side graphs present the median and 75th percentile institutions, respectively.

Unlike R&D expenditure, facility space is a fixed cost capital investment and also a potential capacity constraint for research. Thus, facility space is mostly constant across time for most fields and institution types. How much facility space is “enough”? Does facility space construction signal higher future research output? In the analysis of Section 7, there is a clear positive association between facility space and research expenditure, but there is no causal evidence or analysis of lagged effects.

There are large differences between R1 facility space and R2, HSI, and HBCU facility space across nearly all fields at the median and 75th percentile. This is unsurprising for fields in the life sciences (i.e., biological and health sciences), engineering, and physical sciences. However,
this is even true for fields with lower baseline levels of facility space such as computer science (see Figure 59),\textsuperscript{35} mathematics (see Figure 65),\textsuperscript{36} psychology (see Figure 67), and social sciences (see Figure 68).

The 75th percentile of HSIs had nearly the same facility space as R1s in geosciences and psychology in 2015–2017, but declined again in 2021, as seen in Figures 61 and 67. Figure 62 shows that the median R1 and HBCU both increased facility space in agricultural sciences significantly in the last 10 years. The median HBCU and HSI have more agricultural sciences space than the median R2.

**Facility Cost Over Time**

Since costs are too sparse by field, Figures 70 and 71 show the total R&R costs and new construction costs over time for R1s, R2s, HSIs, and HBCUs.

Figure 70 shows that both the median and 75th percentile R1s have significantly higher R&R costs than other institution types, with a peak in 2017 and steady decrease since then. HSIs and R2s have similar levels of R&R costs, though HSIs are marginally higher. HBCUs have the lowest R&R costs (which are close to R2 R&R costs in 2021) that are relatively constant over time.

New construction costs in Figure 71 are much more volatile, with the median R1 having the highest costs since 2017. The median and 75th percentile R1s and HSIs have historically overlapped and switched places, though in 2021 HSIs have lower new construction costs. R2s and HBCUs have historically similar levels of new construction costs, with R2s having slightly higher costs in 2021.

**KEY TAKEAWAYS**

This section summarizes the major takeaways from examining the time trends of R&D expenditure and facility space in the last decade:

\textsuperscript{35} There is also a substantial increase in computer science facility space of the 75th percentile R1 in the last 10 years.

\textsuperscript{36} The median R1 has less than 4,000 square feet of facility space for mathematics.
1. **Balanced Panel:** Changes over time in the NSF surveys could be due to the entry and exit of institutions. However, the panels are well balanced across institution types, especially in the HERD survey and 2015 onward in the S&E Research Facilities survey.

2. **R&D Expenditure Over Time:** R1s spent significantly more on R&D than all other institution types in all fields. The median R1 increased total spending significantly in the last decade, while spending for other institution types have remained relatively constant. Capitalized equipment expenditure trends mirror that of total R&D expenditure with much lower dollar amounts and more volatility.

   - The median R2, HSI, and HBCU spent less than $10 million over time for any field. On the other hand, the median R1 spends significantly for life sciences ($150 million), engineering ($50 million), and physical sciences ($20 million) while also increasing spending over time.
   - Computer science spending is relatively lower at the baseline ($8 million for the median R1), but both R1 and R2 investment has increased over time, indicating a potential high-growth research field.
   - Mathematics and statistics spending is the lowest across all fields, along with psychology and social sciences, though median R1 expenditure in social sciences has doubled from $5 to $10 million in the last decade.
   - The median R1 decreased spending in geosciences, with the median TCU experiencing a spike in R&D spending in 2019 to $800,000.

3. **Facility Space Over Time:** R1s have significantly more facility space compared with R2s, HBCUs, and HSIs across all fields and at both the median and 75th percentile. Fields with lower baseline levels of facility space are computer science, mathematics, psychology, and social sciences.

   - At the 75th percentile, HSIs and R1s had similar levels of facility space in geosciences and psychology from 2015 to 2017. Both the median R1 and HBCU increased agricultural sciences facility space significantly over the last decade.
4. **Facility Costs Over Time:** R1 R&R costs peaked in 2017 and have since decreased (though they are still the largest). HSIs and R2s have similar levels of R&R costs, while HBCUs have the lowest. New construction costs are much more volatile, with R1s/HSIs and R2s/HBCUs historically having similar levels of costs.
Conclusion

This paper provides a high-level descriptive analysis of the current state of research capacity at MIs versus R1s/R2s across different fields. First, there is an analysis of the current and future state of MI Carnegie Classifications, with an emphasis on the new 2025 classification. Second, there is an analysis of R&D expenditure and facility space, with a focus on the association between the two measures of research capacity. Note that the paper provides suggestive evidence of building up physical infrastructure as an actionable policy, but does not have the analysis to speak to other important determinants of research output, such as researcher and faculty selection. The following sections summarize the conclusions from the two main areas of analysis and suggest next steps for future research.

CARNEGIE CLASSIFICATIONS

There are currently 14 HBCUs classified as R2 (11) and R3 (3) and 41 HSIs classified as R1 (11), R2 (12), and R3 (15). TCU’s are not part of the CCIHE. There is large heterogeneity in HSIs, as supported by the long right tail in research capacity reflected in the rest of this paper. Since HSIs are threshold based, there are many large R1 schools that are classified as HSIs. Currently, there are no HBCUs that are R1, but under the new 2025 classifications that will use clear thresholds for R&D expenditure and research doctorates, Howard University will be the first HBCU to be classified as R1 based on 2021 data.\footnote{Howard University was an R1 in 2000 but was demoted in 2005 after a classification scheme change.} Given the classification change, what is the tangible impact on research capacity and output? Will being classified as an R1 increase funding opportunities, faculty recruitment, or student enrollment? This is a key question for future research and policy consideration.
Research capacity includes R&D and equipment expenditure, as well as facility space and costs. The analysis is based on the HERD and S&E Research Facilities surveys from the NSF, which are the most comprehensive public sources of data on university research capacity. However, the NSF surveys come with R&D thresholds that select on institutions with the most R&D spending as opposed to a census. While the HERD survey includes all R1s and R2s, only half of HBCUs and a quarter of HSIs are included. The S&E Research Facilities survey has even less coverage, with only 60 percent and 75 percent of R1s and R2s included, respectively. There is only one TCU, 40 percent of HBCUs and 8 percent of HSIs. Despite stating a clear threshold, there seems to be sample selection in the S&E Research Facilities survey unattributed to R&D expenditure. The interpretation of analysis should instead be framed as “For minority institutions with relatively higher baseline research capacity, how do they compare to R1s/R2s and what are fields of potential investment?”

Conditional on the sample selection, there is still immense right-skew in the distribution of R&D expenditure and facility space. This is not surprising given the presence of power laws and accumulation effects of research funding. However, this does make summarizing the data difficult, so this paper presents distribution moments such as the median and 75th percentile, along with right-tail institutions (i.e., top 10) when appropriate. From the field-specific analysis of research capacity, there are several key takeaways:

1. **Investment-Heavy Fields:** Life sciences (by a large margin), engineering, and physical sciences (in that order) require intense investment in R&D funding, equipment, and facility space. There is also a high association between facility space and R&D spending for these three fields controlling for graduate student enrollment. On the one hand, this may prove difficult for many MIs to compete with R1s. On the other hand, this may also show areas that are important for research (hence the R&D funding) for MIs to invest in. A possible avenue for growth is to have strategic partnerships between MIs and R1s for fields that require large capital investments.

---

38 There are only five TCUs.
2. **Investment-Light Fields**: Mathematics and statistics (by far), psychology, and social sciences require much less investment in R&D funding, equipment, and facility space. This may be an area for MIs to grow research capacity, though the question then arises of how to attract researchers and what drives the research output in these fields if not physical capital. The association between facility space and R&D expenditure is the weakest for mathematics, and moderately strong for psychology and social sciences (though not at the 25th percentile of R&D spending in the overall sample).

3. **Fields on the Rise**: Computer science is on the rise for R1 spending, especially in equipment and facility space. This is not surprising given the increasing importance of computer science in the last two decades. Geosciences and agricultural sciences also show much less differential between R1 and MI spending. The association between facility space and computer science expenditure is the strongest across all fields, which proves to be promising as a policy tool. The association between facility space and geosciences expenditure is moderate.

When considering where to place funding, the objective of each institution arises to the forefront. The top 10 HBCUs tend to be more specialized than the top 10 HSIs in both R&D spending and facility space, which may call for different strategies of increasing research investment. Should HBCUs focus on increasing research capacity in a few fields, or should they diversify their research portfolio? The association between facility space and R&D spending is also lower for HBCUs across many fields (i.e., computer science, mathematics and statistics, physical sciences, psychology and social sciences), where R&D spending is below the expected average of the whole matched sample given the amount of facility space. The question that arises that cannot be answered by this paper is where research funding would best go. Is it toward building infrastructure, or is it toward attracting researchers and supporting faculty?

On the other hand, there are more than 300 HSIs with a much more heterogeneous distribution that includes 11 R1s. Should smaller HSIs focus on specializing, or should research funding be diverted to the most research-intensive HSIs with the largest student body? How much are MIs competing with each other versus all research institutions for funding and students? TCU's are a special case, with only one institution in the S&E Research Facilities survey and thus
severely lacks data. There is also a community-driven mission, and it is important to determine to what extent expanding research capacity across all fields versus specializing in a few fields is the goal.

**NEXT STEPS**

This paper takes a step toward describing the state of research capacity at minority institutions, but there are many data limitations that prevent a more comprehensive analysis. The following are some next steps for data acquisition:

- The NSF surveys select for MIs with higher levels of R&D spending. The data is left truncated if the goal is to understand the distribution of research at all HBCUs, HSIs, and TCUs. On the other hand, if the goal is to understand research at MIs with already some baseline research capacity, then the NSF surveys suffice.

- This paper provides a preliminary analysis of the distances between MIs and R1s based on data from IPEDS. However, there is much room to explore the geographic spillovers and possible partnerships between research facilities and MIs. Additional data on locations of federally funded R&D centers and national labs, along with possible field specializations, may indicate potential avenues of collaboration.

- Research depends significantly on researchers themselves, such as faculty and students. Though not present in this paper, the NSF provides data on graduate students and postdoctoral researchers. This could be important to merge with the current data to understand the relationship between research spending, capital, and personnel.

- There is no analysis of downstream data on research output of institutions. Are there publication records, faculty movement, patenting, or government procurement deals? Are there other sources of funding that institutions can tap into? These would require merging and cleaning external sources.

This paper has also raised questions that are important to consider when deciding where and how to allocate future research funding to best enhance MI research capacity:
• What is the objective of each institution in terms of research capacity? Is the goal to improve specific fields, or diversify research output across many fields? Is the goal to attract all students and compete with R1s or to focus on the needs of community? Taking into consideration the peer group and objective function of each institution will be instrumental to how policy is implemented. How important is an institution’s Carnegie Classification to subsequent research funding and output? This will help determine the impact of the new 2025 CCIHEs.

• What is the causal relationship between increasing facility space and R&D spending? This paper provides suggestive evidence for fields where the association is stronger; however, this is not causal evidence. The association is also heterogeneous between institution types, where HBCUs have a lower association across many fields compared with HSIs. Why is this the case, and what are the other determinants of R&D spending (e.g., personnel, equipment)? Given causal evidence, funding can then be allocated toward building infrastructure.

• Does colocation of MIs and R1s increase the research capacity of MIs or do they compete in terms of researchers and students? Can institutions create partnerships for equipment and capital sharing (especially in investment-heavy fields)?
References


