How many researchers are there in the world?

Davut Emrah Ayan, Institute for Policy & Social Research, The University of Kansas, Lawrence, KS USA
Laurel L. Haak* (corresponding author), Ronin Institute and Mighty Red Barn, Townsend, WI USA
Donna K. Ginther, Department of Economics and Institute for Policy & Social Research, The University of Kansas, Lawrence, KS USA

Conflict of interest: The authors declare no conflicts of interest.

Funding: LLH was funded by NBER and The Wellcome Trust, NBER Project No. 25410.00.00.00.6600. DKG was funded by NSF SMA-1854849.

Keywords: Innovation capacity, researcher density, GERD, tertiary educational engagement

Abstract

Many studies of innovation focus on expenditures for research and development (R&D). In this report, we use R&D personnel data from international sources to determine global and regional employer research-years, and couple with employee surveys, educational attainment, and population data to estimate global and regional researcher headcount. We argue that the employee survey approach provides a useful method for assessing innovation capacity, and, if adopted more broadly, can provide a strategic framework for countries and regions to develop and support human capital to support innovative activities. We consider the role of funding and R&D personnel in the production of patents and publications and find that R&D personnel measures have more explanatory power.

Note: Slides for NBER meeting are here.

Introduction

At first glance, knowing the number of researchers in the world may seem just another statistic. However, it forces us to examine how we define "researcher", which exposes assumptions about training, degree, discipline, employment sector, and primary outputs. UNESCO reports 7.8 million full-time equivalent (FTE) researchers in 2013 - 0.1% of the global population.[1] This figure represents firm-reported employment data from G20 countries, only 10% of all countries in the world. There are university programs graduating doctorates in over 160 countries (UNESCO), and ranked universities in over 100 countries (Shanghai), suggesting that many more researchers are not included in the UNESCO figure.

If researcher FTEs were just a number, then perhaps we could be satisfied with 7.8 million. However, researcher density has policy significance. Research and development (R&D) investments do not directly translate into innovation capacity.[2] Firm-reported data help us to understand research intensity, but it does not provide a window into work occupations or educational backgrounds of the people performing the R&D. Organizations absorb and utilize knowledge through structural, human, and social capital.[3] Human capital is a key factor in innovating in response to public needs.[4] If we can know the number of researchers in a place, we can infer R&D sector capacity in that place and also better understand knowledge sharing on a local and global scale, in private, public, and government sectors.[5] In turn, this can provide a strategic framework for nations to develop and support human capital for innovative activities, necessary for solving the world's challenging sustainable development goals.

Research and development definitions and data collection methods have evolved. UNESCO involvement in R&D statistical gathering has helped to enable comparisons across countries on a global scale.[6] Initially focused on science and engineering, R&D data collection principles underwent a substantial revision between 2010-2015. R&D is now defined as "creative work

---

[1] UNESCO Science Report (2013). Towards 2030: Facts and Figures, Human Resources. https://en.unesco.org/node/252277 [Accessed 6 September 2021].

[2] Reviewed in Hamdan A and Hamdan R (2020). The mediating role of oil returns in relationship between investment in higher education and economic growth: Evidence from Saudi Arabia. Economics and Sociology, 13(1), 116-131. https://doi.org/10.14254/2071-789X.2020/13-1/8

[3] Zhu H, Zhao S, Abbas A (2020) Relationship between R&D grants, R&D investment, and innovation performance: The moderating effect of absorptive capacity. Journal of Public Affairs 20(1): 14723891. https://doi.org/10.1002/pa.1973; Engelman RM, Fracasso EM, Schmidt S, and Zen AC (2017). "Intellectual capital, absorptive capacity and product innovation", *Management Decision*, 55(3): 474-490. https://doi-org.www2.lib.ku.edu/10.1108/MD-05-2016-0315; Soo C, Tian AW, Teo STT. and Cordery J (2017). Intellectual Capital–Enhancing HR, Absorptive Capacity, and Innovation. Human Resource Management, 56: 431-454. https://doi.org/10.1002/hrm.21783

[4] Belmonte da Silva R and Fernandez Jardón CM (2021) The relationship between human intellectual capital and innovation capacity in the public sector. Visión de Futuro 25(2): 137 -153 https://doi.org/10.36995/j.visiondefuturo.2021.25.02R.004.en; Lewis JM, Ricard LM, Klijn EH (2017). How innovation drivers, networking and leadership shape public sector innovation capacity. International Review of Administrative Sciences 84(2): 268-307. https://doi.org/10.1177/0020852317694085

[5] Wagner CS, Park HW, Leydesdorff L (2015) The Continuing Growth of Global Cooperation Networks in Research: A Conundrum for National Governments. PLoS ONE 10(7): e0131816. https://doi.org/10.1371/journal.pone.0131816; Kristjánsson B, Helms R, Brinkkemper S (2014). Integration by communication: knowledge exchange in global outsourcing of product software development. *Expert Systems* 31(3):267-281. https://doi.org/10.1111/exsy.640

[6] UNESCO (2014). Guide to Conducting an R&D Survey: For Countries Starting to Measure Research and Experimental Development. http://dx.doi.org/10.15220/978-92-9189-151-1-en

undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications."[7] This change, encompassing social sciences and humanities, traditional knowledge, as well as new data coding and collection advice to developing countries, opens opportunities for new analysis of interactions between human resources, investments, and innovation economies on a broader scale than has been afforded by the data prior to 2015.

The traditional approach to comparing R&D capacity across countries has been to compare R&D expenditures. The U.S. National Science Board's Science & Engineering Indicators[8] measures R&D intensity as the ratio of GERD to GDP. No similar metric has been developed when it comes to research personnel. In this paper, we seek to develop and test models of intersecting human and structural factors contributing to R&D sector growth to answer the following key questions: (1) how should we measure how many researchers there are in the world, and (2) is researcher density a useful variable in models of innovation capacity? (3) does researcher density have explanatory power for models of patents and publications? We took a two-stage approach to these objectives. First, we developed a measure of researcher density. Then we proposed an innovation capacity model and tested variable interactions with researcher density. We used data from open sources with global reach and inter-country data quality standards and have created a study data set that is available for re-use.

Methods and Data

We had two primary goals in this study. First was determining how to define and then measure researchers on a global scale. In this paper, we define researchers as those people doing R&D as a primary or secondary job activity (more on this in the following sections). Second was assessing the innovation capacity of countries and regions, by examining relationships between researcher counts (FTEs, counts, and extrapolations) and innovation inputs (country-level investments in R&D, educational engagement and attainment), outputs (research papers and patents), as well as environment (governance metrics and university rankings).

Sources

The data used in our analyses were collected from publicly available data sources, in most cases with global scope. We obtained data on educational statistics, R&D employment, Gross Domestic Expenditure on R&D and general country demographic data (including population) from the World Bank World Development Indicators, UNESCO Institute for Statistics, and OECD; employment data from the International Labour Organization (ILO), and population data from . We also utilized US-specific data on graduate students, publication, and employment,

---

[7] OECD (2015) Frascati Manual: Guidelines for Collecting and Reporting Data on Research and Experimental Development, the Measurement of Scientific, Technologic, and Innovation Activities. OECD Publishing, Paris. https://dx.doi.org/10.1787/9789264239012-en.
[8] National Science Board. 2020. *Science and Engineering Indicators 2020: The State of U.S. Science and Engineering*. NSB-2020-1. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsb20201/; see: Global R&D Intensity definition at https://www.ncses.nsf.gov/pubs/nsb20201/global-r-d#intensity.[Accessed 28 Oct 2021]

namely the US National Science Foundation (NSF) Survey of Doctoral Recipients and the National Survey of College Graduates. Although data collected may share the same sources, data coverage varies by country (see coverage tables in **Appendix 1**).

Downloaded raw data files were cleaned and transformed using Stata/SE version 16.1. Further transformation into data tables and descriptive analysis including graphing and correlation analysis were performed using R version 4.1.0. Multivariable regression analysis was performed using Stata. We will provide the dataset and descriptive metadata via a dedicated landing page.

Maps were created using ESRI's ArcMap (v10.8.1) utilizing country boundaries from ESRI (v10.2, 2015) in the Winkel Tripel projection. Data are grouped into five classes using either the Jenks method or by quintile distribution.

Level of observations in our analysis are the countries and their aggregate statistics for the years from 2014 to 2018. For US-specific analyses, we used data from the 2003 SESTAT and the 2015 SDR (which is matched to publications) and 2017 NSCG. Naturally, countries vary by many dimensions, primarily by population and geography, and these dimensions have consequences for other variables. For that reason, we normalized educational, investment, and employment variables at the country level by dividing by the population segment age 25-69.

Descriptive Statistics

Regions and Countries

Our intent was to maximize the global coverage of our dataset, with a sample size of at least ten countries in each of six regions.[9] From an initial list of 217 countries, we selected those with at least one year of data for the period 2014-2018, for the measures of: (a) doctoral education enrollment or attainment; (b) gross domestic expenditure on research and development (GERD); and (c) researcher full-time equivalents (see "Variables" section below). We expanded the dataset by imputing researcher data when possible.[10]

Some countries that have strong tertiary education sectors did not have GERD or researcher data in the UNESCO or World Bank sources during 2014-2018 (or in the 5 years prior), and could not be included. This was a particular challenge in the Middle East and Africa region, affecting Saudi Arabia, Kenya, and Nigeria. UNESCO partners with the African Science, Technology and Innovation Indicators (ASTII) Initiative of the African Union to support economic and education statistical data collection. This region is home to 12 countries ranked in the top 25 fragile states,[11] four of which (Chad, Ethiopia, Mali, and Mozambique) have sufficient data

[9] We examined international country lists provided by OECD, G20, and World Bank, and used World Bank regional groupings as they provided the most complete coverage. https://data.worldbank.org/country. More details here: https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries

[10] We imputed Australian researcher FTE from Employment in professional, science, and technical activities data; and Peruvian and Israeli researcher FTE from researcher head count data (UNESCO).

[11] The Fund for Peace (2021). Fragility in the World 2021. https://fragilestatesindex.org/ [Accessed 19 Sept 2021].

coverage for inclusion in this study. Our final data set included 105 countries and was augmented with data on population and employment (**Figure 1**).

Region	**Countries**
APAC: East Asia, Pacific, and South Asia (19)	Australia, Brunei Darussalam, Cambodia, China, Hong Kong SAR (China), India, Indonesia, Japan, Macao SAR (China), Malaysia, Myanmar, New Zealand, Pakistan, Philippines, Singapore, South Korea, Sri Lanka, Thailand, Vietnam
ECA: Europe and Central Asia, not EU or Schengen Area (11)	Bosnia and Herzegovina, Georgia, Kazakhstan, Moldova, Montenegro, North Macedonia, Russian Federation, Serbia, Turkey, Ukraine, Uzbekistan
EUS: European Union, Schengen Area, and Common Travel Area (31)	Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, United Kingdom
LATAM: Latin America and Caribbean (15)	Argentina, Brazil, Chile, Columbia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, México, Panamá, Paraguay, Perú, Puerto Rico, Trinidad and Tobago
MEA: Middle East and Africa (27)	Bahrain, Burkina Faso, Burundi, Chad, Dem Rep. Congo, Côte d'Ivoire, Egypt, Ethiopia, Iran, Israel, Jordan, Kuwait, Lesotho, Madagascar, Mali, Mauritius, Mozambique, Namibia, Oman, Qatar, Rwanda, Senegal, Seychelles, South Africa, Togo, Tunisia, United Arab Emirates
NA: North America (2)	Canada, United States

Table 1. Countries in the data set, organized by region.

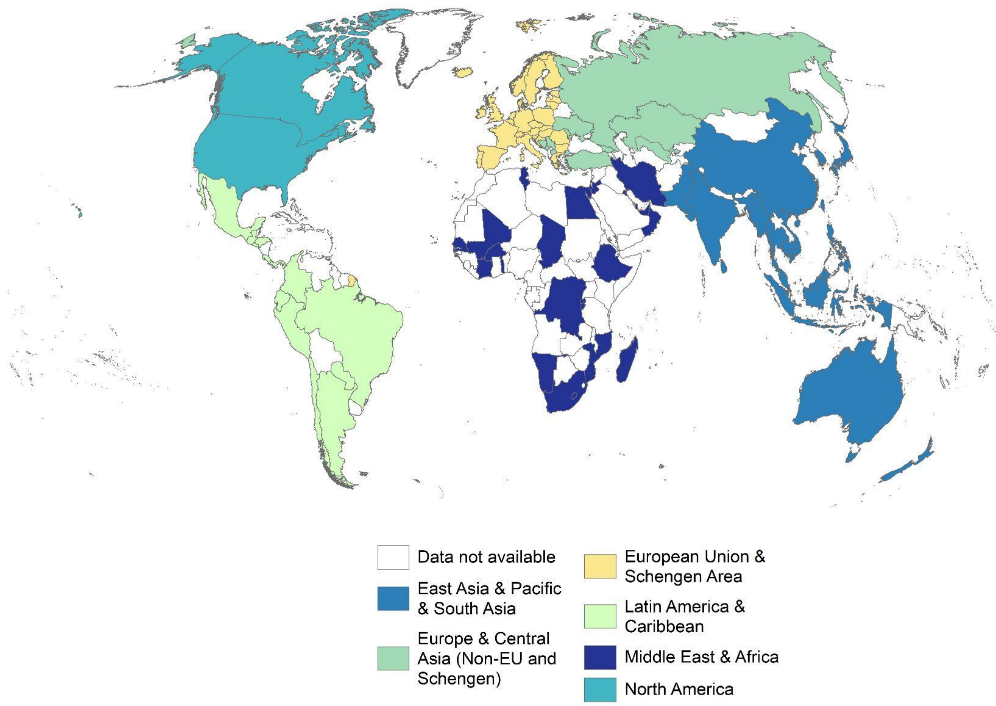

Figure 1. Countries in the final data set shown on a Winkel global projection.

Variables and Coverage

Summary tables of variables and availability for each country, aggregated by region, are provided in **Appendix 1**.

National investments

Country-level data on Gross Domestic Product (GDP) (World Bank World Development Indicators) and Gross Domestic Expenditure on Research and Development (GERD) (UNESCO Institute for Statistics - Science Technology and Innovation 9.5.1) were available for all of the countries in our dataset. For most countries, we were also able to obtain information on government investments in education and business sectors. All investment data were normalized per capita for the 25-69 population subset and log-transformed prior to running correlations and regressions. **Table A1.a** shows country-level data availability for the investment variables GDP, GERD total and by sector, aggregated by region.

Educational Intensity

As noted above, we selected countries based on availability of core educational data. We made the assumption that researchers will have completed a college degree. Ideally, we would have liked to examine the full range of educational enrollments and attainments, from a technical degree (International Standard Classification of Education (ISCED) 5) to a doctoral-level degree

(ISCED 8)[12], as well as field of study, however the data did not support this broad examination for more than OECD countries. We were able to obtain or impute total tertiary (ISCED 5-8) and doctorate (ISCED 8) enrollment or attainment data for all countries in the dataset (UNESCO Institute for Statistics). All educational data were normalized per capita for the age range 25-69 and log-transformed prior to running correlations and regressions. **Table A1.b** shows country-level data availability for the educational variables enrollment and attainment, total tertiary and for doctorates, aggregated by region.

Researcher Counts

The lines between education, research, development, design, and application are difficult to ascertain.[13] We tested several methods to assess how many people are engaged in R&D activities. Our goal was regional comparisons so we focussed on UNESCO data, but we also extrapolated counts from US employment data. **Table A1.c** shows country-level data availability for employment and researcher-related variables, aggregated by region. For regressions and correlations, we normalized all data by per capita for the age range 25-69 and then log transformed the data.

(1) **The total personnel counts and FTEs (full-time equivalents) employed in R&D** (UNESCO Institute for Statistics). These data are derived from country-level employer surveys carried out in business, education, government, and nonprofit sectors. Notably, while business R&D FTE data are available for the US,[14] total R&D FTE data are missing; we imputed total FTE using sector employment ratios from the US National Survey of College Graduates (NSCG), sponsored by the NSF and carried out every other year by the US Census Bureau.[15]

(2) Self-reported data on R&D activity. We were interested in comparing self-reported and employer reported data on R&D activities. While the US does not collect workforce totals from employer surveys, it does field demographic surveys and collect extensive workforce information directly from individuals. These data provide a lens into work activities, occupations, and educational background. We used the NSCG to collect information on survey respondents who reported R&D as a primary and/or secondary activity, as well as educational and occupational variables (**Table A4.b**). We compared

---

[12] The ISCED levels were re-defined in 2011 and implemented in 2014. See UNESCO (2014). Guide to Conducting an R&D Survey: For Countries Starting to Measure Research and Experimental Development. http://dx.doi.org/10.15220/978-92-9189-151-1-en

[13] See examples and discussion in OECD (2015) Frascati Manual: Guidelines for Collecting and Reporting Data on Research and Experimental Development, the Measurement of Scientific, Technologic, and Innovation Activities. OECD Publishing, Paris. https://dx.doi.org/10.1787/9789264239012-en.

[14] US R&D personnel data in the UNESCO database are extrapolated from data provided by the National Science Foundation Center on Science and Engineering Statistics (NCSES) to OECD about 20 years ago. At that time, The NCSES baseline totals for the Government sector were based on the US Office of Personnel Management (OPM) estimates for Federal scientists and engineers with R&D as their primary work activity The NCSES estimates for the Education and Nonprofit sectors were estimated from the NSF Survey of Doctorate Recipients who reported R&D as their primary work activity. The NCSES is fielding new establishment surveys and data for business, government, education, and non-profit sectors will become available over the next 5 years.

[15] Burke A, Finamore J, Foley D, Jankowski J, Moris F; National Center for Science and Engineering Statistics (NCSES). 2021. *Measuring R&D Workers Using NCSES Statistics.* NSF 21-335. Alexandria, VA: National Science Foundation. Available at https://ncses.nsf.gov/pubs/nsf21335/.

these figures with those from employer survey data after normalizing for total tertiary attainment, and adjusted this ratio based on country-level reporting variations for FTE and counts (see below) to estimate R&D personnel counts.

(3) **The total number of people who completed a college degree** (ISCED 5-8) (UNESCO Institute for Statistics). This provides an estimate of tertiary educational engagement and a broad scope of country-level reservoir for innovation, but may not include people who contribute to R&D through traditional knowledge, self-taught, or trades pathways.

(4) **The total number of people who completed a doctoral degree** (ISCED 8) (UNESCO Institute for Statistics). This provides a more focused estimate of educational engagement for those people who have engaged in a course of study that encourages novel thinking and research approaches. We extended this analysis using NSF data to explore relationships between doctorates, publications, and grants.

Outputs

We focused on data that were indicative of innovative activity[16] that were also available for our country data set. We used patenting activity (patent application by residents) from the World Intellectual Property Organization Patent Report and research publication volume from the National Science Foundation, Science and Engineering Indicators. To mitigate skewing and the adverse effects of outliers, we binned and log-transformed each variable to obtain a more normal distribution for correlation and regression analyses. We captured the count of ranked universities per country, using the 2019 Shanghai Academic Ranking of World Universities and 2020 CWTS Leiden Ranking which presents data for the time period 2015-2018; these surveys are largely based on faculty productivity as measured by paper production. We used total counts normalized per million population, and also binned these data (0, 1-4, 5 or more). **Table A1.d** shows country-level output data availability for patent applications, publications, and ranked universities, aggregated by region.

Social Factors

In addition to economic and educational factors, we also examined the impact of social factors on innovation capacity. For this, we used the World Bank Worldwide Governance Indicators data on government effectiveness, control of corruption, political stability, rule of law, voice and accountability, and regulatory quality compiled from over 30 sources reporting the perceptions of governance of a large number of survey respondents and expert assessments. Data were available for all of the countries in our dataset worldwide. We used index scores for each country, averaging across the 2014-2018 year range. **Table A1.e** shows country-level governance environment data availability, aggregated by region.

[16] OECD/Eurostat (2018), *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition*, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris/Eurostat, Luxembourg, https://doi.org/10.1787/9789264304604-en.

Summary Statistics

Summary tables are presented in **Appendix 2,** and include counts, and when applicable, mean, median, and standard deviation for each variable. We also report p-values from the test of equality of means across regions (ANOVA). We did not assume equal variance, and before we tested equality of means, we tested homogeneity of variances (Bartlett test in R) and for all variables homogeneity of variances were rejected at the 5% significance level. North America is excluded from equality of means analysis because of its small sample size (n=2).

Exploratory Data Analysis

We first examined the distribution of data for each variable. Descriptive statistics and histograms show that data are positively skewed even after normalization by population. To mitigate the effect of skewness and outliers in our data, we log transformed variable data. Software packages R and Stata use natural logarithm by default for log transformation.

Correlation Analysis

We investigated how our variables are related to each other, especially with our outcome variables, by graphing scatter plots and calculating correlation coefficients. Although these results may not translate into causal relationships, it is important to see the direction and strength of bivariate relationships before developing multivariate models. In Appendix 3, **Table A3.a-d**, we provide Pearson correlation coefficients indicating the strength of the linear correlation for variable pairs for all primary variables: R&D investments, tertiary education engagement, R&D human capital, and innovation indicators. We find significant correlations for all pairs except for ranked universities and patents.

Multivariate Model Specifications

In a multivariate regression setting, we model our outcome variables as a function of education, investment, R&D intensity, R&D personnel intensity, and governance variables. We estimate our empirical models by Ordinary Least Squares as in (1) below. For models with log-log specification, coefficients can be interpreted as elasticities.

$$\boldsymbol{Y_i} = \boldsymbol{\alpha} + \boldsymbol{\beta X_i} + \boldsymbol{\delta Z_i} + \boldsymbol{\epsilon_i} \qquad (1)$$

Our analysis uses country level data for one time period (2014-2018). $\boldsymbol{Y_i}$ is the outcome in each model for each country, $\boldsymbol{X_i}$ is a matrix of education, employment and investment variables if included in the models; $\boldsymbol{Z_i}$ is a matrix of governance variables. Each model also includes region dummies in order to account for region-specific heterogeneity. Regression analysis is performed using Stata/SE version 16.1.

Results

Counting Researchers

Global Data: Employer Surveys

We started the counting process by examining UNESCO UIS R&D personnel data. These data are collected using an annual employer questionnaire[17] based on definitions of R&D and personnel encoded in the Frascati Manual:[18]

- R&D is defined to "comprise creative and systematic work undertaken in order to increase the stock of knowledge – including knowledge of humankind, culture and society – and to devise new applications of available knowledge." R&D includes basic and applied research and experimental development.

- R&D personnel includes all persons engaged directly in R&D, whether directly employed or external contributors, as well as those providing direct services for the R&D activities such as R&D managers, administrators, researchers, technicians, and clerical staff. R&D personnel perform scientific and technical work for an R&D project (setting up and carrying out experiments or surveys, building prototypes, etc.); plan and manage R&D projects; prepare interim and final reports for R&D projects; provide internal services for R&D projects (e.g. dedicated computing or library and documentation work), and provide support for the administration of the financial and personnel aspects of R&D projects.

- R&D personnel excludes individuals undertaking indirect support or ancillary activities in R&D-performing units, such as specific services to R&D provided by central computer departments and libraries, services by central finance and personnel departments dealing with R&D projects and R&D personnel, and the provision of services for security, cleaning, maintenance, canteens, etc., to R&D-performing units.

Figure 2 shows total R&D FTEs using the UNESCO data for all countries in our study dataset. We include, for context, researcher FTEs (a subset of R&D FTEs), doctorate and all tertiary education counts. Not all R&D human resources have a doctorate, and some proportion do not have a tertiary degree. The exact degree of overlap differs by country and in many cases is not evident in the data.

[17] Questionnaires are available for download on the UNESCO UIS Website at http://www.uis.unesco.org/UISQuestionnaires/Pages/country.aspx. [Accessed 25 September 2021].
[18] OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264239012-en

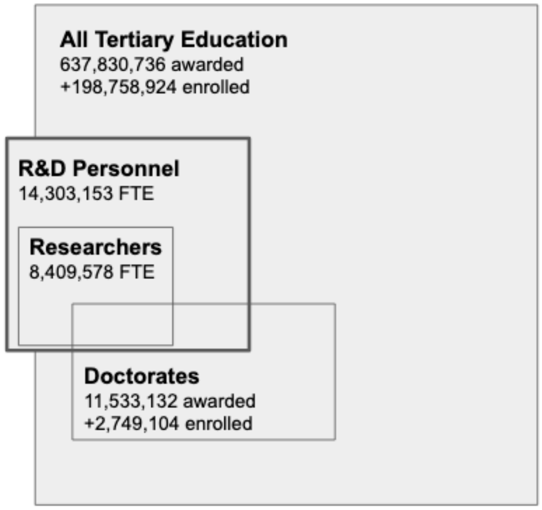

Figure 2. R&D Human Resources, average 2014-2018. SOURCE: UNESCO.

Extrapolations: Individual Surveys

As with any dataset covering multiple countries and variables, there are vagaries.[19] As noted above in the Methods section, UNESCO R&D human resource data for the US is an extrapolation from a 20-year old baseline. Further, the UNESCO data are derived from employer-based surveys and do not provide an individual perspective on research activities. Given the shift to a more inclusive definition of R&D in 2015, the employer-based approach likely does not capture some types of research activities and hence undercounts people engaged in R&D activities.

To address these issues and explore other perspectives on R&D activities, we turned to the National Survey of College Graduates (NSCG), a biennial survey conducted by the US Census Bureau. The NSCG asks respondents whether they spend 10% of their time or more each week on basic research, applied research or development. Respondents are asked to choose their primary and secondary work activities. According to NSCG weighted tabulations, there were 3.432m US college graduates engaged in R&D (defined as a work activity that is basic research, applied research, or development) as a primary work activity, 6.054m college graduates engaged in R&D as a secondary work activity in 2017, and an additional 11.28m individuals who report spending at least 10% of their time at work on some aspect of R&D. We created an algorithm to convert NSCG headcount data to estimated FTE to enable a first-pass comparison between NSCG and UNESCO R&D personnel figures, and found a 2.6-fold higher FTE count based on the NSCG data (**Table 2**).[20] Over 20 million people in the US report some R&D work activity and we estimate 5.65 million R&D FTE personnel.

[19] Brunei Darussalam, Columbia, Costa Rica, and Cote d'Ivoire report only Researcher FTE. Their total FTE reported is less than 0.1% of the total FTE for all countries in our data set. We have imputed R&D personnel data from researcher FTE or count data, when available.
[20] We assigned weights based on full-time and part-time employment, primary, secondary and any R&D work activity.

Source	Variable	Effort	Headcounts	Weights	Estimated FTE
UNESCO UIS	US R&D Personnel FTE				2,163,950
US Census NSCG[21]	Counts, primary work activity is R&D	Full time	2,987,628	1	2,987,628
		Part time	444,636	0.5	222,318
	Counts, secondary work activity is R&D	Full time	5,118,129	0.25	1,279,532
		Part time	936,101	0.1	93,610
	People whose primary or secondary work activity is *not* R&D, BUT who report their work involves R&D	Full time	10,079,617	0.1	1,007,962
		Part time	1,202,259	0.05	60,113
	Totals		*20,768,370*		*5,651,163*

Table 2. US R&D Human Resources, 2014-2018. SOURCE: NSCG (2017) and UNESCO (2014-2018).

Over half of people indicating research as a primary work activity have a Bachelor's terminal degree, and two-thirds are employed in the business sector. Engineering, Computer and Mathematical Sciences, General Management, Biological and Agriculture and Other Life Sciences, and Physical and Life Sciences are the top 5 occupations represented, making up 67.0% of the total. Writers, Editors, Press, and Historians are also well-represented, with 6.8% of the total. Additional demographics for the NSCG sample are shown in **Table A4.** Clearly, R&D activities are performed by a broad spectrum of talent, across sectors, and in a variety of occupations, although it should be noted that the NSCG does not capture those individuals performing research who do not have a college degree.

As a thought experiment, we extrapolated from NSCG data for personnel reporting R&D as a primary activity, to estimate global R&D personnel counts (**Table 3**). We applied to each country R&D FTE value (from UNESCO) the ratio: US *R&D FTE / US tertiary education attainment per capita*, then normalized by tertiary education attainment per capita for each country. FTEs are fractionals of total personnel counts; we further transformed the figures using country-level FTE/headcount ratios (from UNESCO data) to obtain an estimate of R&D persons. Our extrapolation suggests that R&D FTE personnel in the world may be 62% greater than reported by UNESCO.

[21] Adjusted to remove indirect or ancillary activities, as specified in the Frascati definition.

Region	R&D FTE (UNESCO)	Extrapolated R&D FTE	Extrapolated R&D Count
APAC	6,580,119	10,470,991	17,365,522
ECA	1,081,415	2,017,061	2,578,329
EUS	3,129,929	3,765,298	5,721,266
LATAM	590,423	1,807,345	2,938,128
MEA	522,539	1,235,678	2,166,905
NA	2,398,728	3,896,502	4,267,428
Totals	*14,303,153*	*23,192,875*	*35,037,579*

Table 3. Researcher FTEs and extrapolated FTEs and headcounts, by region and totals.

If we go further, and include in our estimates not only those who report R&D as a primary work activity, but also those who report R&D as a secondary activity, the global total rises to 97 million.

Doctorates as R&D Personnel

Another way of assessing innovation is to focus specifically on the doctorate population and examine the relationship between self-reported R&D work activities, publication authorship, and government research support. Our goal is to examine whether authorship or research grants are proxies for being a researcher. For this analysis, we used the US National Science Foundation Survey of Doctoral Recipients (SDR) from 2015. NSF has linked respondents in its 2015 SDR[22] to Web of Science publications from 1990-2017. SDR respondents also report whether their work currently receives US Federal government research support (**Table 4**).

	Weighted total	Number with at least one publication	Number with at least one US federal grant
Number of doctorates reporting primary work activity is R&D	364,337	311,579	130,327
Number of doctorates reporting secondary work activity is R&D	301,539	246,944	78,583
Total number of doctorate reporting employed status	787,250	732,439	216,328

Table 4. Relationship between doctorate status, authorship, grants, and R&D occupation. SOURCE: NSGG and SDR.

[22] National Science Foundation, Division of Science Resources Statistics, *Characteristics of Doctoral Scientists and Engineers in the United States: 2003*, NSF 06-320, Project Officer, John Tsapogas (Arlington, VA 2006).

Of the employed respondents, about 85% were in occupations that were either primary or secondary R&D focussed. Furthermore, over 80% of doctorates employed in R&D occupations are linked to at least one publication, strongly suggesting that authorship can be used as a proxy for researcher status, at least among individuals with doctorate degrees. **Figure 3a** shows the overlap between primary work activity, secondary work activity and being an author on at least one publication. Only 14% of those reporting primary work as R&D and only 18% of those reporting secondary work as R&D do not have publications. It should be noted, however, that attempts to divine the number of unique authors based on disambiguation of publication datasets are fraught by issues with name ambiguity,[23] low coverage of non-English language journals and disciplinary variations in publication venues by field.[24]

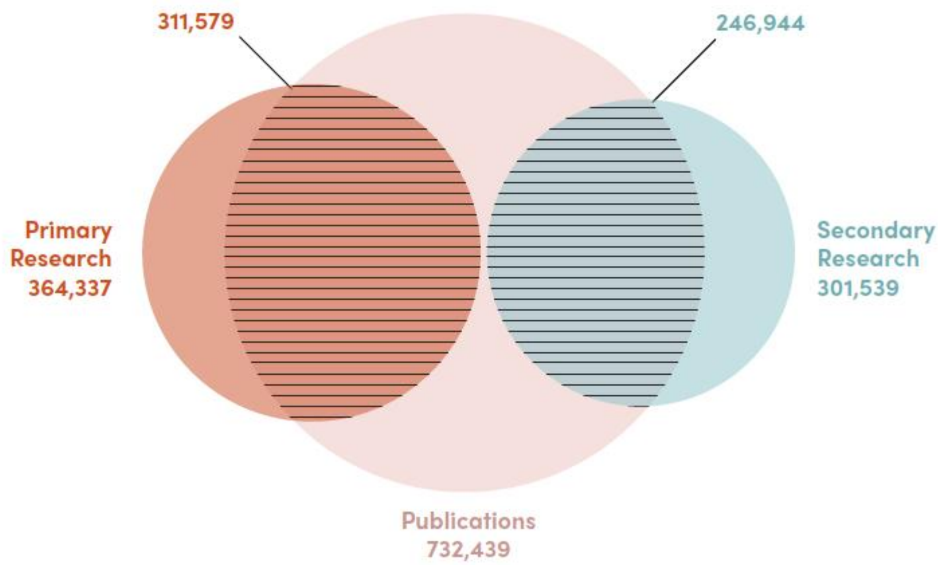

Figure 3a. Publications. Proportion of US Doctorates with at least one publication, 1990-2017. SOURCE: 2015 NSF Survey of Doctorate Recipients matched to Web of Science publications.

Federal funding was less associated with doctorate employment, with about 27% of employed doctorates reporting having received any government research support in 2015. This may be misleading, as it excludes non-federal awards from foundations, industry seed grants, and similar. **Figure 3b** shows the overlap between government support and primary and secondary R&D work activities. Government research support is a poor measure of R&D activities. Only 36% of those whose primary work activity and 26% of those whose secondary work activity is R&D have government support.

---

[23] Kim J (2019) Scale-free collaboration networks: An author name disambiguation perspective. *Journal of the Association for Information Science & Technology*. 70(7):685-700. https://doi.org/10.1002/asi.24158

[24] Bello M and Galindo-Rueda F (2020). Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2). *OECD Science, Technology and Industry Working Papers*, No. 2020/03, OECD Publishing, Paris, https://doi.org/10.1787/1b06c47c-en.

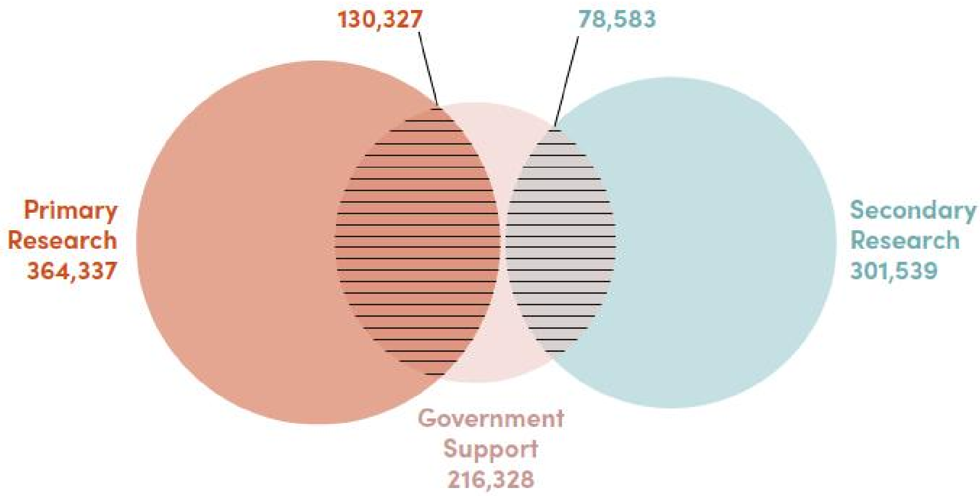

Figure 3b. Government Support. Proportion of US Doctorates with government research support. SOURCE: NSF Survey of Doctorate Recipients, 2015.

Visualizing Research Intensity and Researcher Density

Because our objective is not just to count researchers, but also to assess differences between modes of measuring *researcher density*, we mapped normalized data to ascertain qualitative differences between R&D FTEs (**Figure 4**) and R&D persons (**Figure 5**)[25]; we include a map of the standard *research intensity* innovation metric GERD as a share of GDP for comparison (**Figure 6**). Comparing these maps, there are clear differences that may help to reshape our understanding of the innovation capacity of countries.

First, examining modes of measurement, we see the highest researcher densities in Australia , irrespective of mode. When comparing R&D FTEs (figure 4) to extrapolated R&D personnel headcounts (Figure 5), we see qualitative shifts in density, with relative regional increases in Latin America, Asia Pacific, and Central Asia; mixed effects in African and North American countries, and decreases in Europe.

Comparing research intensity (Figure 6) to researcher density (Figure 4), we see different patterns. There is relatively more R&D spending vs. personnel in China, India and Russia; and higher R&D personnel vs. spending In Australia and New Zealand; in Kazakhstan and Uzbekistan; across Europe; in Morocco; in Costa Rica, Chile, and Argentina; and in the US.

These measures - research intensity and researcher density - present different aspects of a country's research capacity and should be used together to assess innovation capacity.

[25] Actual UNESCO headcount data is shown in Appendix 5.

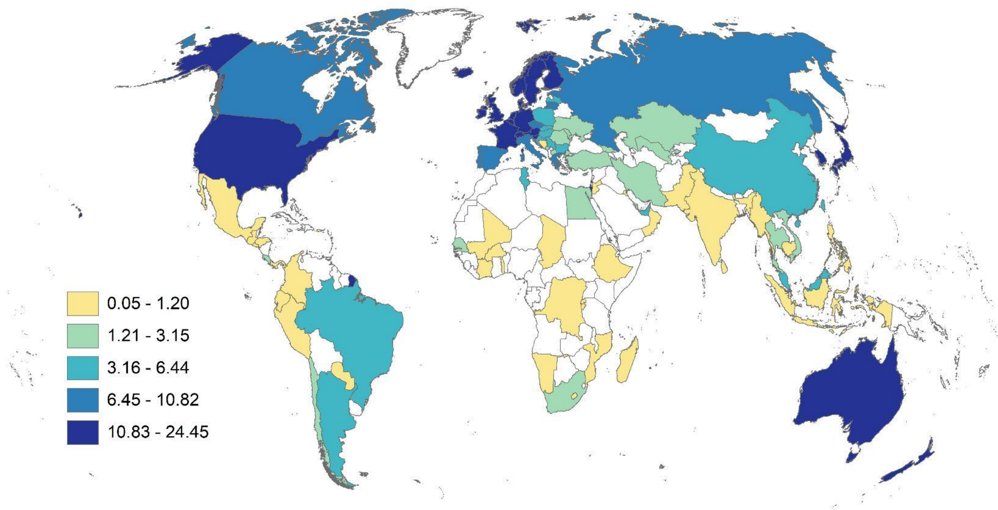

Figure 4. R&D personnel FTE, per 1000 population ages 25-69, Jenks method. Data derived from UNESCO and World Bank sources.

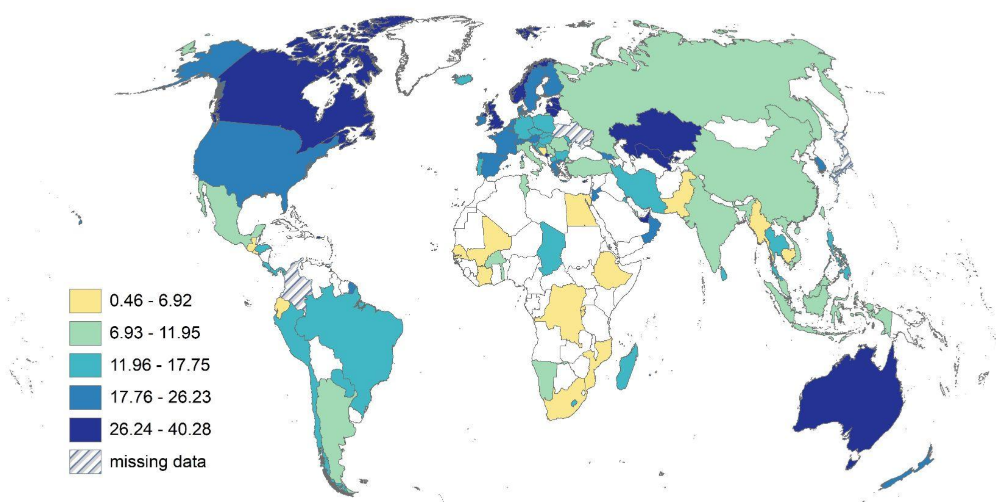

Figure 5. R&D person headcounts, per 1000 population ages 25-69, Jenks distribution. Data extrapolated from UNESCO and NSCG sources.

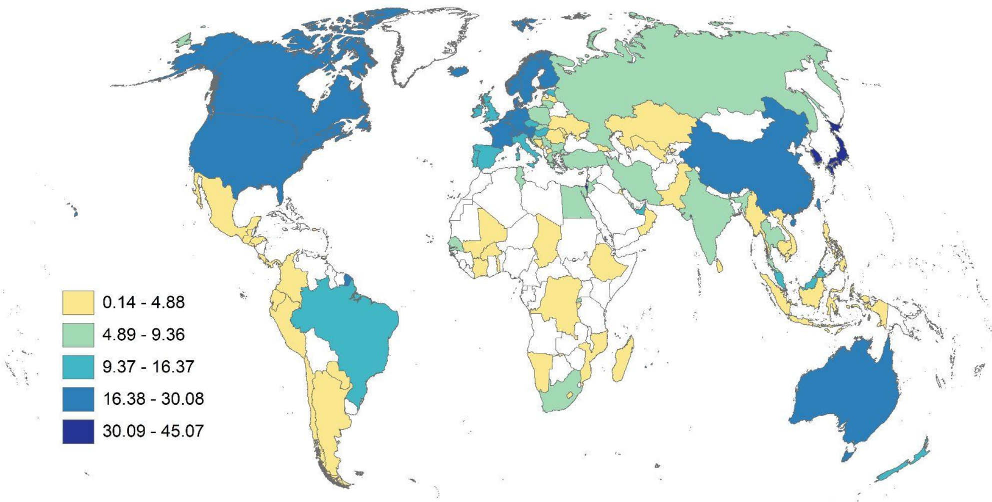

Figure 6. GERD as a proportion of GDP, per 1000 population ages 25-69, Jenks distribution. Data derived from UNESCO and World Bank sources.

Regressions

What is the relative importance of R&D personnel and R&D funding in the production of research output measured by publications and patents? To explore the association between researchers and research output measured by publications and patents, we estimated a multivariate regression model where measures of researchers and research funding were included as covariates to test a series of models, and we also examined interactions at the regional level using scatter plot analysis.

Using our cross section of national data, we estimate models of the impact of R&D personnel and different types of GERD (total, business, and higher education). Our model also includes controls for GDP, educational enrollment and attainment, total ranked universities, as well as variables for governance and regional dummies (**Table A6.a**). The log of R&D FTE personnel are adjusted per capita and regressed on the log of patent applications creating elasticity measures. A 1% increase in R&D personnel was associated with a .89 to 1% increase in patent applications depending on the measure of GERD used in the models. None of the estimates of GERD were associated with increased patenting. We performed the same thought experiment with log publications. A 1% increase in R&D personnel was associated with approximately .85% increase in publications. In the publication model, tertiary enrollment per capita was associated with increased publications as was GERD higher education funding. Interestingly, there was not a significant effect of ranked universities on publications.

We probed these results further in **Table A6.b**. In our first model, we include the log of GERD as a share of GDP, the measure of research intensity. In the second model we include the log of R&D personnel per capita, the measure of researcher density, and in the last model we include both. A 1% increase in research intensity is associated with a 1.34% increase in patent applications and .82% increase in publications. A 1% increase in researcher density is associated with .96% more patent applications and .87% more publications. When we include both research intensity and researcher density in the models, the coefficients on GERD as a share of GDP drop in magnitude and are no longer statistically significant. However, researcher density remains significantly associated with increases in both patents and publications. These analyses point to future research directions for developing researcher density policy frameworks.

Discussion

In this paper, we have described an alternative way of measuring innovation capacity, using researcher density. We have used this approach to estimate global counts of people doing R&D work. We were inspired by the UNESCO operating definition of research and development, a broad definition that encompasses traditional knowledge, humanities and social sciences, product design, engineering, and sciences. Indeed, R&D is not defined by field, occupation, or education, but rather, as an "activity [that is] novel, creative, uncertain in its outcome, systematic and transferable and/or reproducible."

At present, UNESCO uses firm-based (employer) surveys to gather the volume of research performed by employees. This is interesting because it indicates the current research activity of an organization, and can provide insight on R&D activities at the national level. However, it does not shed light on the occupations or educational background of the people engaged in research activities, which would help in recruiting and workforce policy decisions.

To get this kind of information requires surveying individuals directly. One such survey, carried out in the US since the 1970s, is the National Survey of College Graduates. While it assuredly misses some people engaged in R&D who don't have college degrees, it does provide a window into the work experiences of college graduates in the US workforce, providing information that connects work activities with occupational data and tertiary educational attainment.

Looking at the R&D workforce from these two perspectives yields substantially different numbers. Firm-based surveys show about 14m R&D FTEs in the world, while extrapolating from the NSCG data suggests the number may be as high as 97m. This discrepancy surely has an impact on government educational and employment policies, as well as firm-level recruiting plans such use of visa programs, for example. It should also have bearing on curriculum, training, and professional development expectations for undergraduate and graduate students.

Further, our findings suggest that individual-level data on occupation and R&D activities - or, researcher density - may also be used to develop country-level measures of capacity for innovation. This measure shows the number of people who are actively engaged in research, and who can be recruited into innovative projects as needed to solve tough problems. In fact our regression analysis demonstrates that researcher density is significantly associated with publications and patenting. Research intensity--GERD as a share of GDP-- no longer has explanatory power after including measures of researcher density.

We argue that the policy question is not measuring innovation as an end in itself, but rather the ability to engage in innovative processes as a means to solve problems as they arise. Better understanding researcher density, as well as interactions with other economic indicators such as GERD, tertiary education enrollment and attainment, ranked universities, and governance environment will be necessary for effective policy development. Some innovative approaches to assessing workforce show that non-governmental data sources can be very useful in measuring workforce mobility,[26] wage interactions with educational institution,[27] as well as the impact of academic freedom on innovation.[28] At the same time, UNESCO data collection of R&D variables has improved global coverage and makes it possible to perform regional and world-wide studies, which are necessary for cross-national work toward sustainable development goals, and we encourage UNESCO to consider including researcher density approaches in their R&D collection processes.

[26] Gomez CJ, Herman AC, Paolo Parigi P (2020). Moving more, but closer: Mapping the growing regionalization of global scientific mobility using ORCID. Journal of Informetrics, 14(3):101044. https://doi.org/10.1016/j.joi.2020.101044.

[27] Martinelli P, Schoellman T, Sockin (2021) Alma Mater Matters: College Quality, Talent, and Development. Available at SSRN: https://ssrn.com/abstract=3899337 or http://dx.doi.org/10.2139/ssrn.3899337

[28] Berggren N and Bjørnskov (2021). Academic Freedom, Institutions, and Productivity. IFN Working Paper No 1405. Available at SSRN: https://ssrn.com/abstract=3927675

Appendix 1. Data Coverage Tables

Variable Name [Source]	APAC (19)	EUS (31)	ECA (11)	LATAM (15)	MEA (27)	NA (2)	Total (105)
GDP [WDI]	19	31	11	15	27	2	105
GERD [UNESCO]	19	31	11	15	27	2	105
GERD - Business Sector [UNESCO]	16	31	10	11	15	2	85
GERD - Education Sector [UNESCO]	19	31	11	14	25	2	102

Table A1.a. Countries in the dataset with investment data coverage for at least one year 2014-2018. Data sources are indicated in [square brackets]. Total country count per region is indicated in (parentheses).

Variable Name [Source]	APAC (19)	EUS (31)	ECA (11)	LATAM (15)	MEA (27)	NA (2)	Total (105)
Enrollment ISCED 5 [UNESCO]	19	26	6	12	23	2	88
Enrollment ISCED 6 [UNESCO]	19	31	10	12	27	2	101
Enrollment ISCED 7 [UNESCO]	19	31	10	12	27	2	101
Enrollment ISCED 8 [UNESCO]	19	31	11	13	27	2	103
Total Tertiary Enrollment ISCED 5-8 [UNESCO]	19	31	11	13	27	2	103
Attainment ISCED 5-8 {imputed} [UNESCO]	11 {18}	29 {31}	7 {10}	12 {14}	15 {27}	2	76{102}
Attainment ISCED 6-8 {imputed} [UNESCO]	9 {18}	26 {30}	6 {8}	11 {14}	15 {27}	2	69{98}
Attainment ISCED 7-8 {imputed} [UNESCO]	10 {18}	25 {31}	6 {10}	10 {14}	15 {27}	2	68{102}
Attainment ISCED 8 {imputed} [UNESCO]	9 {18}	25 {31}	5 {10}	6 {14}	14 {27}	2	61{102}

Table A1.b. Countries in the dataset with education data coverage for at least one year 2014-2018. Data sources are indicated in [square brackets]. Total country count per region is indicated in (parentheses). Imputed data are indicated by {curly brackets}.

Variable Name [Source]	APAC (19)	EUS (31)	ECA (11)	LATAM (15)	MEA (27)	NA (2)	Total (105)
Total Employment [ILO]	19	31	11	15	26	2	104
Total Employment - Education Sector [UNESCO]	4	26	2	5	1	0	38
R&D FTE - Total [UNESCO]	17 {19}	31	11	11 {15}	23 {27}	1 {1}	94{105}
R&D FTE - Business Sector [UNESCO]	16	31	9	7	16	2	81
R&D FTE - Education Sector [UNESCO]	18	31	10	10	23	1	93
R&D FTE - Government Sector [UNESCO]	17	31	10	10	22	1	91
Researcher FTE [UNESCO]	17{19}	31	11	13{15}	23{27}	2	97{105}
Technician FTE [UNESCO]	16{17}	20	11	11	23	1	82{83}
R&D Counts - Total [UNESCO]	17 {19}	31	11	14	26	0 {2}	99{101}
R&D Counts - Business Sector [UNESCO]	14	31	10	7	18	1	81
R&D Counts - Education Sector [UNESCO]	17	31	11	13	25	0	97
R&D Counts - Government Sector [UNESCO]	15	31	11	13	25	0	95

Table A1.c. Employment and researcher data coverage for at least one year 2014-2018. Data sources are indicated in [square brackets]. Total country count per region is indicated in (parentheses). Imputed data are indicated by {curly brackets}.

Variable Name [Source]	APAC (19)	EUS (31)	ECA (11)	LATAM (15)	MEA (27)	NA (2)	Total (105)
Patent applications - residents [WIPO]	18	31	10	13	17	2	91
Research publications [NSF]	17	31	11	15	27	2	103
Ranked University in Country >0 [Shanghai or Leiden]	13	29	3	5	10	2	62

Total Ranked Universities in Region [Shanghai or Leiden]	429	394	44	45	69	236	1217

Table A1.d. Ranked universities and patent and publication data coverage in 2019. Data sources are indicated in [square brackets]. Total country count per region is indicated in (parentheses).

Variable Name [Source]	**APAC (19)**	**EUS (31)**	**ECA (11)**	**LATAM (15)**	**MEA (27)**	**NA (2)**	**Total (105)**
Government Effectiveness [WGI]	19	31	11	15	27	2	105
Control of Corruption [WGI]	19	31	11	15	27	2	105
Political Stability [WGI]	19	31	11	15	27	2	105
Rule of Law [WGI]	19	31	11	15	27	2	105
Voice and Accountability [WGI]	19	31	11	15	27	2	105
Regulatory Quality [WGI]	19	31	11	15	27	2	105

Table A1.e. Governance data coverage for at least one year 2014-2018. Data sources are indicated in [square brackets]. Total country count per region is indicated in (parentheses).

Appendix 2. Overall Summary Statistics

We collected statistics from sources as described in the Methods section, averaged by country (**Table A2.a**) and region (**Tables A2.b-h**) over the time period 2014-2018. The sample showed variation by region in the income groups represented, however, employment per capita was not significantly different across regions (**Table A2.b**).

We examined patent applications and research publication volume, as well as the number of ranked universities, by country, for the time period 2014-2018 (**Table A2.c).** There is substantial variation in each variable by region. APAC, EUS, and NA have a 10-fold higher count of ranked universities. NA and APAC had 30-fold higher patent volumes than other regions but overall there was not a significant regional variation (ANOVA, p=0.233). NA publication volume was 5-80 times higher than other regions, and EUS had a moderately higher count of publications but not patents; overall there was a significant regional variation in publication outputs (ANOVA, p=0.0176**). Each of these variables is positively skewed, evident in the difference between the mean and median, as well as the large standard deviation. For correlation and regression analysis, we log-transformed patent and publication variables, and binned the university counts.

We examined economic factors that may influence innovation. In addition to total country-level GDP, we collected data on gross expenditures on research and development (GERD), as well as sub-categories of GERD for the business and education sectors (**Table A2.d**). Regional data were, for the most part, normally distributed, and showed 2-10 fold differences between regions (excluding NA). Overall, there was highly significant variation between regions for all GERD variables (ANOVA, p<0.0001, ***). We used per capita normalized variables for correlations and regressions.

Human resources play an important role in innovation. We collected data on people employed in research and development occupations (**Table A2.e**), as well as people enrolled in tertiary education (**Table A2.f**) or with higher education degrees (**Table A2.g)** -- groups that include R&D personnel and potential to engage in R&D activities. While absolute personnel FTEs varied substantially between countries and regions, this variation was reduced when normalized by population. This effect was more pronounced for tertiary educational enrollment and attainment variables. A number of countries in our sample were missing data for attainment during the 2014-2018 timeframe, and we imputed values based on available education data for these countries from 2010-2013. Overall, there was significant regional variation in R&D personnel (ANOVA, p=0.0254, **) and researcher variables (ANOVA, p= 0.411, **), and highly significant variability for all education variables (ANOVA, p<0.001, ***) We used population-normalized variables and imputed attainment variables for correlations and regressions.

Governance may also impact research and development activities. We collected country-level governance scores for each country, averaged the scores across the time period 2014-2018, and created regional summary statistics (**Table A2.h**). The scores varied significantly across regions (ANOVA, p<0.0001, ***). We used average scores in correlation and regression analyses.

Variable	N	Mean	Std. Dev.	Min	Pctl. 50	Max
Region	105					
... East Asia & Pacific & South Asia	19	18.10%				
... Europe & Central Asia (Non-EU and Schengen)	11	10.50%				
... European Union & Schengen Area	31	29.50%				
... Latin America & Caribbean	15	14.30%				
... Middle East & Africa	27	25.70%				
... North America	2	1.90%				
Number of ranked universities	105	11.59	31.75	0	1.00	226.00
Number of Patent applications, residents	91	22,353	124,593	1.75	378	1,122,778
Number of Scientific and technical journal articles	103	22,481	63,950	12.42	2,782	447,684
GDP - per capita	105	48,369	37,738	2,431	43,088	188,064
GERD - per capita	105	503.10	700.61	0.50	167.39	2,782.64
GERD Business Sector - per capita	85	379.07	539.07	0.05	120.34	2,411.37
GERD Higher Education Sector - per capita	102	133.29	173.38	0.13	51.86	705.10
R&D personnel FTE - per 1000 population	105	5.07	5.43	0.05	2.64	24.45
Researcher FTE - per 1000 population	105	3.40	3.72	0.03	1.62	15.58
Employment - per capita	104	0.83	0.18	0.49	0.82	1.37
Tertiary attainment, total - per 1000 population	102	251.28	149.81	8.39	224.96	775.30
Tertiary attainment, Doctorate - per 1000 population	102	4.96	5.73	0.11	2.58	35.50
Tertiary enrollment, total - per 1000 population	103	60.49	29.43	9.98	58.88	155.53
Tertiary enrollment, Doctorate - per 1000 population	103	1.34	1.35	0.01	0.93	6.08
Government Effectiveness	105	0.41	0.92	(1.58)	0.24	2.22
Control of Corruption	105	0.28	0.99	(1.40)	0.02	2.25
Political stability and Absence of Violence	105	0.06	0.88	(2.41)	0.13	1.53
Regulatory Quality	105	0.45	0.92	(1.46)	0.42	2.18
Rule of Law	105	0.33	0.96	(1.62)	0.20	2.05
Average score on "Voice and Accountability"	105	0.20	0.93	(1.83)	0.24	1.69

Table A2.a. Summary statistics for the study data set.

Variable →	N	Income Group [UNESCO]				Employment, per capita [ILO]	
Region		**High**	**Upper Middle**	**Lower Middle**	**Low**	**Mean (Median)**	**SD**
APAC	19	8	4	7	0	0.87 (0.88)	0.12
ECA	11	0	8	3	0	0.72 (0.77)	0.13
EUS	31	30	1	0	0	0.77 (0.76)	0.09
LATAM	15	4	2	9	0	0.87 (0.88)	0.14
MEA	27	8	4	5	10	0.91 (0.87)	0.26
NA	2	2	0	0	0	0.85	

Table A2.b. Income distribution of countries in sample, by region. SOURCE: UNESCO and ILO

	Patent applications, residents [WIPO]			Number of Scientific and technical journal articles [NSF]			Ranked Universities [Shanghai, CWTS]				
	N	**Mean (Median)**	**SD**	**N**	**Mean (Median)**	**SD**	**Region Total**	**Ave (SD)**	**0**	**1-4**	**5+**
APAC	18	87,224 (1,094)	267,675	17	50,419 (10,094)	108,468	429	22.6 (52.0)	6	4	9
ECA	10	3,614 (272)	7,958	11	10,288 (610)	19,854	44	4 (9.24)	8	1	2
EUS	31	3,517 (811)	8,977	31	21,491 (11.173)	28,754	394	12.7 (17.0)	2	10	19
LATAM	13	612 (53)	1,372	15	6,420 (583)	14,361	45	3 (7.69)	10	2	3
MEA	17	1,016 (16)	3,355	27	3,456 (223)	8,594	69	2.6 (7.12)	17	6	4
NA	2	146,871		2	244,701		236	118	0	0	2

Table A2.c. Patent application, publications, and ranked university counts, by region. SOURCES: WIPO, NSF, Shanghai, and CWTS.

	GERD, per capita			GERD % of GDP			GERD Business Sector, per capita			GERD Education Sector, per capita		
	N	**Mean (Median)**	**SD**	**N**	**Mean (Median)**	**SD**	**N**	**Mean (median)**	**SD**	**N**	**Mean (Median)**	**SD**
APAC	19	576.7 (232.7)	787.4	19	1.06 (0.68)	1.17	16	422.9 (191.1)	600.5	19	150.6 (30.05)	188.2
ECA	11	101.1 (60.78)	97.13	11	0.47 (0.36)	0.33	10	48.46 (19.27)	58.89	11	31.42 (24.88)	32.54
EUS	31	965.4 (790.2)	699.9	31	1.68 (1.36)	0.88	31	611.5 (467.2)	493.4	31	250.3 (193.6)	185.7

LATAM	15	88.66 (59.19)	83.57	15	0.32 (0.27)	0.31	11	33.86 (27.10)	43.26	14	24.00 (13.71)	22.62
MEA	27	223.4 (72.76)	557.6	27	0.55 (0.35)	0.84	15	224.0 (14.90)	623.21	25	61.52 (13.77)	125.12
NA	2	1,734		2	2.22		2	1,141		2	377.8	

Table A2.d. Gross expenditure in research and development (GERD), total and by sector. Per capita is normalized to the total population from ages 24-69 years. GDP is expressed as PPP, in constant 2017 international dollars. SOURCES: UNESCO (GERD) and WDI (GDP, population).

	R&D Personnel, FTE			**Researchers, FTE**			**R&D Personnel, FTE per 1000 capita**			**Researchers, FTE per 1000 capita**		
	N	**Mean (Median)**	**SD**	**N**	**Mean (Median)**	**SD**	**N**	**Mean (Median)**	**SD**	**N**	**Mean (Median)**	**SD**
APAC	18	346,322 (66,396)	903,788	19	188,981 (53,850)	401,211	19	4.86 (2.52)	5.02	18	3.66 (1.79)	3.97
ECA	11	98,310 (17,242)	236,328	11	56,778 (12,821)	126,487	11	2.93 (2.23)	2.23	11	1.94 (1.45)	1.19
EUS	31	100,965 (46,688)	152,499	31	64,400 (33,127)	94,863	31	9.93 (10.27)	4.62	31	6.59 (6.73)	3.21
LATAM	15	39,362 (3,208)	104.063	15	22,019 (1,642)	55,685	15	1.14 (1.12)	0.96	15	0.65 (0.57)	0.68
MEA	27	19,353 (2.588)	38,170	27	12,169 (1,162)	23,835	27	2.22 (0.63)	4.78	27	1.37 (0.38)	3.00
NA	2	1,199,364		2	769,859		2	11.29		2	7.45	

Table A2.e. Research and development human resources, by full-time equivalent per 1000 capita, normalized to the total population from ages 24-69 years. SOURCE: UNESCO

	N	**Total Count (ISCED 5-8)**		**Doctoral Count (ISCED 8)**		**Total per 1000 capita**		**Doctorates per 1000 capita**	
		Mean (Median)	**SD**	**Mean (Median)**	**SD**	**Mean (Median)**	**SD**	**Mean (Median)**	**SD**
APAC	19	5,592,964 (1,745,539)	11,741,985	45,374 (21,160)	79,645	59.46 (51.88)	26.93	1.23 (0.59)	1.35
ECU	11	1,478,984 (250,669)	2,504,964	22,484 (2,511)	39,040	65.02 (62.42)	33.41	0.95 (1.06)	0.80
EUS	31	654,240 (300,576)	834,117	25,309 (14,835)	40,244	67.97 (65.53)	18.54	2.58 (2.41)	1.28
LATAM	13	1,770,855 (705,498)	2,360,030	14,761 (1,012)	29,772	87.74 (79.94)	27.61	0.44 (0.22)	0.43
MEA	27	445,523 (123,587)	938,080	8,609 (1,002)	23,728	35.55 24.82)	23.32	0.51 (0.15)	0.75
NA	2	10,446,042		215,383		88.83		2.23	

Table A2.f. Tertiary education enrollment, by region. Total includes 2-year (ISCED 5) and 4-year post-secondary degrees (ISCED 6), Masters degrees (ISCED 7), and doctoral degrees (ISCED 8), as defined in the Frascati Manual. SOURCE: UNESCO.

	Total Count (ISCED 5-8)			Doctoral Count (ISCED 8)			Total per 1000 capita			Doctorates per 1000 capita		
	N	Mean (Median)	SD	N	Mean (Median)	SD	N	Mean (Median)	SD	N	Mean (Median)	SD
APAC	18 (7)	15,998,016 (5,429,231)	32,612,114	18 (9)	132,758 (33,701)	206,520	18 (7)	228.7 (189.9)	133.5	18 (9)	3.14 (1.36)	3.70
ECA	10 (3)	5,547,149 (932,607)	8,491,168	10 (5)	55,856 (10,185)	100,118	10 (3)	341.1 (269.9)	206.1	10 (5)	2.59 (2.01)	1.62
EUS	31 (2)	3,340,324 (1,447,725)	4,800,708	31 (6)	95,021 (40,462)	167,384	31 (2)	338.5 (354.5)	91.70	31 (6)	8.84 (7.23)	7.14
LATAM	14 (2)	3,550,291 (793,052)	5,290,920	14 (8)	44,595 (15,428)	75,957	14 (2)	182.5 (179.8)	69.88	14 (8)	2.65 (0.99)	2.78
MEA	27 (12)	1,258,614 (445,718)	2,377,282	27 (13)	28,008 (9,009)	41,519	27 (12)	146.6 (143.8)	124.0	27 (13)	2.96 (2.24)	3.67
NA	2 (0)	53,579,142		2 (0)	2,129,364		2 (0)	547.8		2 (0)	16.49	

Table A2.g. Tertiary education attainment, by region. Total includes 2-year (ISCED 5) and 4-year post-secondary degrees (ISCED 6), Masters degrees (ISCED 7), and doctoral degrees (ISCED 8), as defined in the Frascati Manual. Values imputed from earlier years (2010-2014) for the number of countries shown in parentheses in the *N* column. SOURCE: UNESCO

	N	Government Effectiveness		Control of Corruption		Political Stability, Absence of Violence		Regulatory Quality		Rule of Law		Voice and Accountability	
		Mean (Med)	SD	Mean (Med)	SD	Mean (Med)	SD	Mean (Med)	SD	Mean (Med)	SD	Mean (Med)	SD
APAC	19	0.64 (0.40)	0.98	0.32 (-0.28)	1.08	0.09 (0.14)	1.07	0.56 (0.20)	1.06	0.42 (0.01)	1.01	-0.14 (-0.13)	0.91
ECU	11	-0.11 (0.03)	0.37	-0.47 (-0.45)	0.53	-0.50 (-0.30)	0.67	-0.06 (-0.02)	0.63	-0.36 (-0.23)	0.40	-0.41 (-0.14)	0.68
EUS	31	1.17 (1.10)	0.56	1.10 (0.91)	0.81	0.75 (0.77)	0.39	1.21 (1.17)	0.49	1.20 (1.13)	0.63	1.14 (1.18)	0.37
LATAM	13	-0.09 (-0.06)	0.47	-0.31 (-0.46)	0.57	-0.09 (-0.08)	0.48	0.12 (0.10)	0.60	-0.27 (-0.50)	0.64	0.27 (0.24)	0.45
MEA	27	-0.25 (-0.29)	0.82	-0.17 (-0.08)	0.70	-0.51 (-0.51)	0.87	-0.22 (-0.38)	0.72	-0.20 (-0.14)	0.70	-0.52 (-0.63)	0.75
NA	2	1.64		1.63		0.83		1.62		1.71		1.27	

Table A2.h. Country-level governance indicators, average score over 2014-2018. SOURCE: WGI.

Appendix 3. Correlation Statistics

Variables	Log GERD Total per capita	Log GERD Higher Education per capita	Log GERD Business Sector per capita
Log GDP per capita	0.86****	0.83****	0.76****
Log Employment per capita	-0.33***	-0.32**	-0.27*
Government Effectiveness	0.89****	0.85****	0.80****
Control of Corruption	0.80****	0.78****	0.74****
Political stability and Absence of Violence	0.65****	0.66****	0.54****
Regulatory Quality	0.79****	0.78****	0.69****
Rule of Law	0.85****	0.82****	0.77****
Voice and Accountability	0.62****	0.59****	0.57****

Table A3.a. Strength of linear correlations between variable pairs: R&D investments. *** p<0.01, ** p<0.05, * p<0.10.

Variables	Number of ranked universities	Log Tertiary Enrollment, Doctorate - per 1000 pop	Log Tertiary Enrollment, total - per 1000 pop	Log Tertiary attainment, Doctorate - per 1000 pop	Log Tertiary attainment, total - per 1000 pop
Log GDP - per capita	0.17	0.61****	0.60****	0.56****	0.72****
Log Employment - per capita	-0.05	-0.42****	-0.44****	-0.19	-0.34***
Log GERD Total - per capita	0.33***	0.72****	0.55****	0.63****	0.64****
Log GERD Higher Education - per capita	0.23*	0.70****	0.52****	0.64****	0.67****
Log GERD Business Sector per capita	0.32**	0.71****	0.44****	0.70****	0.74****
Government Effectiveness	0.23*	0.67****	0.51****	0.59****	0.65****
Control of Corruption	0.17	0.58****	0.38****	0.60****	0.55****
Political stability and Absence of Violence	0.05	0.43****	0.35***	0.42****	0.55****
Regulatory Quality	0.15	0.64****	0.50****	0.56****	0.61****
Rule of Law	0.19	0.64****	0.45****	0.63****	0.61****
Voice and Accountability	0.08	0.62****	0.52****	0.54****	0.48****

Table A3.b. Strength of linear correlations between variable pairs: ranked universities and engagement with tertiary education. *** p<0.01, ** p<0.05, * p<0.10.

Variables	Log FTE - Researcher - per 1000 pop	Log FTE - R&D personnel - per 1000 pop
Log GDP per capita	0.79****	0.80****
Log Employment per capita	-0.40****	-0.37***
Log GERD Total - per capita	0.91****	0.90****
Log GERD Higher Education - per capita	0.84****	0.82****
Log GERD Business Sector - per capita	0.88****	0.88****
Log Tertiary Enrollment, Total - per 1000 pop	0.54****	0.51****
Log Tertiary Enrollment, Doctorate - per 1000 pop	0.81****	0.80****
Log Tertiary attainment, total - per 1000 pop	0.67****	0.67****
Log Tertiary attainment, Doctorate - per 1000 pop	0.65****	0.67****
Government Effectiveness	0.83****	0.82****
Control of Corruption	0.73****	0.73****
Political stability and Absence of Violence	0.60****	0.61****
Regulatory Quality	0.72****	0.71****
Rule of Law	0.79****	0.78****
Voice and Accountability	0.62****	0.63****

Table A3.c. Strength of linear correlations between variable pairs: R&D human capital. *** p<0.01, ** p<0.05, * p<0.10.

Variables	Log Number of Patent applications, residents	Log Number of Scientific and technical journal articles
Log GDP - per capita	0.25*	0.54****
Log Employment - per capita	-0.25*	-0.35***
Log GERD Total - per capita	0.54****	0.73****
Log GERD Higher Education - per capita	0.37***	0.64****
Log GERD Business Sector - per capita	0.59****	0.70****
Log Tertiary Enrollment, Total - per 1000 pop	0.36***	0.57****
Log Tertiary Enrollment, Doctorate - per 1000 pop	0.49****	0.68****
Log Tertiary attainment, total - per 1000 pop	0.39***	0.47****
Log Tertiary attainment, Doctorate - per 1000 pop	0.41****	0.50****
Log Researcher FTE - per 1000 pop	0.55****	0.70****
Log R&D personnel FTE - per 1000 pop	0.55****	0.70****
Government Effectiveness	0.33**	0.58****
Control of Corruption	0.24*	0.45****
Political stability and Absence of Violence	-0.03	0.22*

Regulatory Quality	0.20	0.47****
Rule of Law	0.28**	0.52****
Voice and Accountability	0.24*	0.40****

Table A3.d. Strength of linear correlations between variable pairs: Innovation indicators. *** p<0.01, ** p<0.05, * p<0.10.

Appendix 4. NSCG Researcher demographics

		R&D as Primary Work Activity	R&D as Primary or Secondary Work Activity
Degree			
	Bachelors (ISCED 6)	59%	58%
	Masters (ISCED 7)	27%	30%
	Doctorate (ISCED 8)	12%	9%
	Professional	2%	3%
Sector			
	Education	17%	22%
	Government	12%	11%
	Business	71%	68%
Occupation			
	Computer and Mathematical Sciences	14.4%	16.6%
	Biological, Agriculture, and Other Life Sciences	8.1%	4.2%
	Physical Sciences	7.1%	2.4%
	Social Sciences	3.6%	3.3%
	Engineering	25.0%	16.5%
	Health Occupations	4.6%	7.3%
	General Management	12.4%	15.1%
	Teachers, K-12	0.7%	6.2%
	Teachers, Postsecondary	2.3%	3.5%
	Social Work	2.2%	2.5%
	Sales and Marketing	3.5%	7.2%
	Writers, Editors, Historians, PR	6.8%	4.3%
	Administrative Services	3.9%	3.5%
	Professional Services	4.3%	4.0%
	Construction, Precision Production, Maintenance, Transportation, and Other Occupations	3.9%	3.4%

Table A4. US R&D Human Resources by Degree, Sector, and Occupation, 2017. SOURCE: NSCG

Appendix 5. Reported R&D Headcounts

In addition to mapping RD FTE and extrapolated headcount data, we also mapped headcount data from UNESCO data. As figure A5.a shows, in addition to several countries with missing data (US, Canada, Columbia, India, Australia), we see qualitative differences across regions, with higher relative headcounts in Latin America and APAC countries, and mixed differences in Europe and Central Asia.

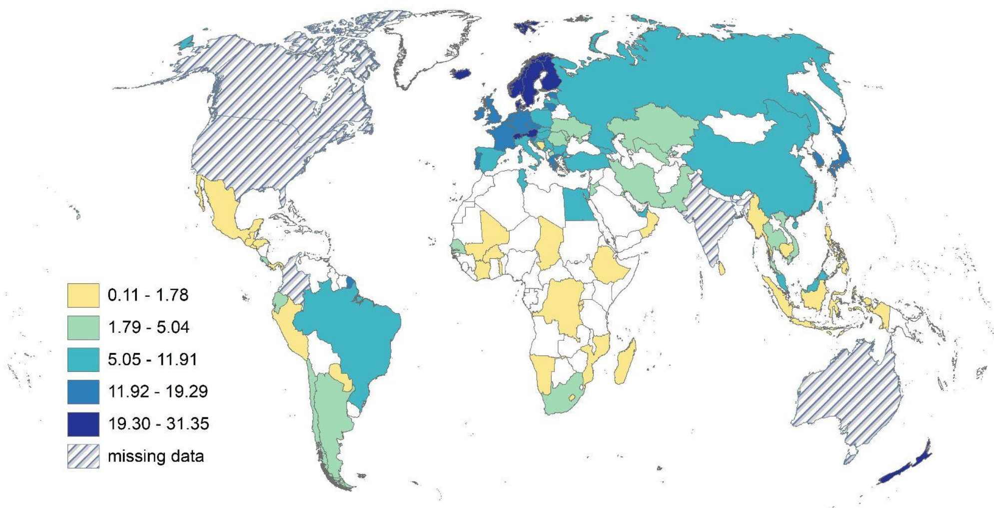

Figure A.5. R&D person headcounts, per 1000 population ages 25-69, Jenks distribution. SOURCE: UNESCO.

Appendix 6. Regression Tables

In all regression tables, robust standard errors are shown in parentheses, and significance is labelled as *** p<0.01, ** p<0.05, * p<0.10.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**Log patent applications**			**Log publications**		
Log R&D personnel FTE, per capita	**0.891*** (0.108)**	**0.997*** (0.105)**	**1.018*** (0.096)**	**0.846*** (0.060)**	**0.832*** (0.053)**	**0.852*** (0.071)**
Log total tertiary enrollment, per capita	0.227 (0.362)	0.275 (0.359)	-0.060 (0.355)	**0.656*** (0.210)**	**0.554** (0.235)**	0.363 (0.238)
Log total tertiary attainment, per capita	0.422 (0.276)	0.443 (0.333)	**0.636* (0.322)**	-0.125 (0.143)	-0.179 (0.144)	-0.117 (0.185)
Log GERD, per capita	0.296 (0.245)			0.092 (0.132)		
Log GERD Higher Education, per capita		-0.073 (0.213)			**0.180* (0.095)**	
Log GERD Business Sector, per capita			0.027 (0.162)			0.051 (0.082)
Log GDP, per capita	-0.447 (0.461)	-0.155 (0.500)	-0.274 (0.389)	0.136 (0.257)	0.095 (0.220)	0.329 (0.307)
Total ranked universities	**0.017*** (0.003)**	**0.017*** (0.003)**	**0.017*** (0.003)**	0.002 (0.002)	**0.003* (0.002)**	0.003 (0.002)
Constant	-4.334 (4.327)	-7.300 (4.470)	-6.193 (4.453)	**-4.354* (2.511)**	-3.296 (2.316)	-4.853 (3.097)
Observations	88	85	79	99	96	80
R-squared	0.900	0.896	0.906	0.943	0.946	0.940

Table A6.a. Patent application and publication volumes vs. economic and education variables. Each model includes governance variables (Government effectiveness, control of corruption, rule of law, political stability, voice and accountability, regularity quality) and region dummies. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**Log patent applications**			**Log publications**		
Log of GERD per 1000 GDP	**1.339*** (0.254)**		0.301 (0.254)	**0.815*** (0.254)**		0.065 (0.129)
Log RD personnel - fte, per 1000 pop		**0.960*** (0.087)	**0.887*** (0.108)		**0.869*** (0.052)	**0.855*** (0.060)
Log total tertiary enrollment, per capita	-0.084 (0.411)	0.291 (0.323)	0.191 (0.353)	**0.829**** (0.352)	**0.713**** (0.190)	**0.704**** (0.201)
Log total tertiary attainment, per capita	**0.899**** (0.273)	0.312 (0.278)	0.387 (0.264)	0.164 (0.260)	-0.087 (0.142)	-0.077 (0.148)
Total ranked universities	**0.032**** (0.006)	**0.017**** (0.003)	**0.017**** (0.003)	0.018 (0.005)	0.002 (0.002)	0.002 (0.002)
Constant	-1.647 (2.362)	-8.047** (2.288)	-7.583** (2.241)	2.240 (1.904)	-3.319** (1.041)	-3.250** (1.063)
Observations	88	88	88	99	99	99
R-squared	0.789	0.897	0.900	0.792	0.942	0.942

Table A6.b. Patent application and publication volumes vs. economic and education variables. Each model includes governance variables (Government effectiveness, control of corruption, rule of law, political stability, voice and accountability, regularity quality) and region dummies. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10.