Appendix: Identifying salary ranges for jobs relevant to the data cycle

Lars Vilhuber

2019-11-12

In the main text, we identify certain job descriptions, with associated salary ranges, from L (lowest) to VH (highest). This appendix identifies possible job titles and associated salary ranges observed in workplace and occupational surveys conducted by the Bureau of Labor Statistics (Bureau of Labor Statistics 2019b).

# Data used

## Occupational Employment Statistics

Collection methods, estimation methodology, and coverage are described in Bureau of Labor Statistics (2019c). We downloaded the data from <https://www.bls.gov/oes/special.requests/oesm18nat.zip> on 2019-10-30. From the downloaded data, we used national\_M2018\_dl.xlsx.

## Occupational Information Network (O\*NET)

The Occupational Information Network (O\*NET) database comprises worker attributes and job characteristics. Information is collected using a two-stage design in which:

* a statistically random sample of businesses expected to employ workers in the targeted occupations is identified and
* a random sample of workers in those occupations within those businesses are selected. Data is collected by surveying job incumbents using a randomly assigned standardized questionnaire on occupation characteristics, out of three questionnaires. Additional questions cover tasks and demographic information.
* Abilities and Skills information is developed by occupational analysts using the updated information from incumbent workers (National Center for O\*NET Development 2019a).

The resulting data covers several thousand occupations. A data dictionary (National Center for O\*NET Development 2019b) provides additional information.

We downloaded version 23\_2 of the data (National Center for O\*NET Development 2019c) from <https://www.onetcenter.org/dl_files/database/db_23_2_excel.zip> on 2019-10-30. From the downloaded data, we used both Occupation Data.xlsx and Alternate Titles.xlsx.

# Methods

## Mapping job titles to SOC

O\*Net is structured around Standard Occupational Categories (SOC) (Bureau of Labor Statistics 2019a). Our main text has a normative list of job description based on data management practiced at university libraries. These may not match reported standard occupation titles exactly. The O\*Net data provides a long but not exhaustive list of alternate mentions of job titles for specific occupations (Alternate Titles.xlsx). Using both the standard occupation title as well as the alternate mention, we match the normative job title via probabilistic matching, using the Jaro-Winkler distance (Winkler 1990) as implemented in the R package fuzzyjoin (Robinson 2019). We keep all *reasonable* matches () to obtain a list of similar occupations and their SOC codes.

## Mapping SOC into salary ranges

OES computes for each SOC code a salary range, comprised of annual salary and hourly wages, and characterized by the 25th and 75th percentile, as well as the median. We attach the annual salary distributions to each of the identified occupations (Table 1), and then collapse these statistics to a triplet of information for each normative job description (Table 2). To do so, we chose to use the minimum of all observed 25th percentiles, the median of all observed medians, and the maximum of all observed 75th percentiles. No weights were applied. An alternative implementation might use the employment shares to create weighted statistics. We do not attempt to compute reliability statistics, as the resulting table is meant to be indicative, not precise. Finally, Table 3 reports

# Results

Table 1 ([nlm.extract.csv](../analysis/nlm.extract.csv)) lists the annual salaries, as of 2018, by job title (median, and the 25% and 75% percentile), for all occupations identified as having similar names as the normative description. Blank salaries (“NA”) indicate that no occupation code could be found on O\*Net based on the normative description.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Job Title | Title | SOC | Alternate Title | A\_PCT25 | A\_MEDIAN | A\_PCT75 |
| Researcher | Industrial Ecologists | 19-2041 | Researcher | 53580 | 71130 | 94590 |
| Researcher | Anthropologists | 19-3091 | Researcher | 48020 | 62410 | 80230 |
| Researcher | Historians | 19-3093 | Researcher | 40670 | 61140 | 85700 |
| Researcher | Biofuels/Biodiesel Technology and Product Development Managers | 11-9041 | Scientist | 112400 | 140760 | 173180 |
| Researcher | Mathematicians | 15-2021 | Scientist | 73490 | 101900 | 126070 |
| Researcher | Chemical Engineers | 17-2041 | Scientist | 81900 | 104910 | 133320 |
| Researcher | Nanosystems Engineers | 17-2199 | Scientist | 69890 | 96980 | 126200 |
| Researcher | Manufacturing Engineering Technologists | 17-3029 | Scientist | 47500 | 63200 | 80670 |
| Researcher | Biologists | 19-1020 | Scientist | 56730 | 77550 | 103540 |
| Researcher | Biochemists and Biophysicists | 19-1021 | Scientist | 64230 | 93280 | 129950 |
| Researcher | Bioinformatics Scientists | 19-1029 | Scientist | 60250 | 79590 | 98040 |
| Researcher | Medical Scientists, Except Epidemiologists | 19-1042 | Scientist | 59580 | 84810 | 118040 |
| Researcher | Chemists | 19-2031 | Scientist | 56290 | 76890 | 103820 |
| Researcher | Hydrologists | 19-2043 | Scientist | 61280 | 79370 | 100090 |
| Researcher | Remote Sensing Scientists and Technologists | 19-2099 | Scientist | 75830 | 107230 | 136930 |
| Researcher | Geographers | 19-3092 | Scientist | 63270 | 80300 | 96980 |
| Data Librarian | Librarians | 25-4021 | NA | 46130 | 59050 | 74740 |
| Data Librarian | Library Science Teachers, Postsecondary | 25-1082 | Librarian | 56550 | 71560 | 90550 |
| Data Librarian | Archivists | 25-4011 | Librarian | 38090 | 52240 | 71250 |
| Metadata Librarian | Librarians | 25-4021 | NA | 46130 | 59050 | 74740 |
| Metadata Librarian | Library Science Teachers, Postsecondary | 25-1082 | Librarian | 56550 | 71560 | 90550 |
| Metadata Librarian | Archivists | 25-4011 | Librarian | 38090 | 52240 | 71250 |
| Records Management Specialist | Librarians | 25-4021 | NA | 46130 | 59050 | 74740 |
| Records Management Specialist | Library Science Teachers, Postsecondary | 25-1082 | Librarian | 56550 | 71560 | 90550 |
| Records Management Specialist | Archivists | 25-4011 | Librarian | 38090 | 52240 | 71250 |
| Curator | Curators | 25-4012 | NA | 39580 | 53780 | 72830 |
| Curator | Archivists | 25-4011 | NA | 38090 | 52240 | 71250 |
| Curator | Archeologists | 19-3091 | Curator | 48020 | 62410 | 80230 |
| Research Domain Curator | Biofuels/Biodiesel Technology and Product Development Managers | 11-9041 | Scientist | 112400 | 140760 | 173180 |
| Research Domain Curator | Mathematicians | 15-2021 | Scientist | 73490 | 101900 | 126070 |
| Research Domain Curator | Chemical Engineers | 17-2041 | Scientist | 81900 | 104910 | 133320 |
| Research Domain Curator | Nanosystems Engineers | 17-2199 | Scientist | 69890 | 96980 | 126200 |
| Research Domain Curator | Manufacturing Engineering Technologists | 17-3029 | Scientist | 47500 | 63200 | 80670 |
| Research Domain Curator | Biologists | 19-1020 | Scientist | 56730 | 77550 | 103540 |
| Research Domain Curator | Biochemists and Biophysicists | 19-1021 | Scientist | 64230 | 93280 | 129950 |
| Research Domain Curator | Bioinformatics Scientists | 19-1029 | Scientist | 60250 | 79590 | 98040 |
| Research Domain Curator | Medical Scientists, Except Epidemiologists | 19-1042 | Scientist | 59580 | 84810 | 118040 |
| Research Domain Curator | Chemists | 19-2031 | Scientist | 56290 | 76890 | 103820 |
| Research Domain Curator | Climate Change Analysts | 19-2041 | Scientist | 53580 | 71130 | 94590 |
| Research Domain Curator | Hydrologists | 19-2043 | Scientist | 61280 | 79370 | 100090 |
| Research Domain Curator | Remote Sensing Scientists and Technologists | 19-2099 | Scientist | 75830 | 107230 | 136930 |
| Research Domain Curator | Anthropologists | 19-3091 | Scientist | 48020 | 62410 | 80230 |
| Research Domain Curator | Geographers | 19-3092 | Scientist | 63270 | 80300 | 96980 |
| Research Domain Project Manager | Biofuels/Biodiesel Technology and Product Development Managers | 11-9041 | Scientist | 112400 | 140760 | 173180 |
| Research Domain Project Manager | Mathematicians | 15-2021 | Scientist | 73490 | 101900 | 126070 |
| Research Domain Project Manager | Chemical Engineers | 17-2041 | Scientist | 81900 | 104910 | 133320 |
| Research Domain Project Manager | Nanosystems Engineers | 17-2199 | Scientist | 69890 | 96980 | 126200 |
| Research Domain Project Manager | Manufacturing Engineering Technologists | 17-3029 | Scientist | 47500 | 63200 | 80670 |
| Research Domain Project Manager | Biologists | 19-1020 | Scientist | 56730 | 77550 | 103540 |
| Research Domain Project Manager | Biochemists and Biophysicists | 19-1021 | Scientist | 64230 | 93280 | 129950 |
| Research Domain Project Manager | Bioinformatics Scientists | 19-1029 | Scientist | 60250 | 79590 | 98040 |
| Research Domain Project Manager | Medical Scientists, Except Epidemiologists | 19-1042 | Scientist | 59580 | 84810 | 118040 |
| Research Domain Project Manager | Chemists | 19-2031 | Scientist | 56290 | 76890 | 103820 |
| Research Domain Project Manager | Climate Change Analysts | 19-2041 | Scientist | 53580 | 71130 | 94590 |
| Research Domain Project Manager | Hydrologists | 19-2043 | Scientist | 61280 | 79370 | 100090 |
| Research Domain Project Manager | Remote Sensing Scientists and Technologists | 19-2099 | Scientist | 75830 | 107230 | 136930 |
| Research Domain Project Manager | Anthropologists | 19-3091 | Scientist | 48020 | 62410 | 80230 |
| Research Domain Project Manager | Geographers | 19-3092 | Scientist | 63270 | 80300 | 96980 |
| Informatician | Computer Systems Analysts | 15-1121 | NA | 68730 | 88740 | 113460 |
| Informatician | Information Technology Project Managers | 15-1199 | IT Specialist | 66410 | 90270 | 117070 |
| Data Wrangler | Information Technology Project Managers | 15-1199 | IT Specialist | 66410 | 90270 | 117070 |
| Education Specialist | Health Educators | 21-1091 | Education Specialist | 39800 | 54220 | 74660 |
| Education Specialist | Special Education Teachers, Secondary School | 25-2054 | Education Specialist | 48630 | 60600 | 77820 |
| Education Specialist | Instructional Coordinators | 25-9031 | Education Specialist | 49280 | 64450 | 82860 |
| Communication Specialist | Public Relations Specialists | 27-3031 | Communication Specialist | 44490 | 60000 | 81550 |
| Software Engineer | Computer and Information Research Scientists | 15-1111 | Software Engineer | 91650 | 118370 | 149470 |
| Software Engineer | Software Developers, Applications | 15-1132 | Software Engineer | 79340 | 103620 | 130460 |
| Software Engineer | Software Developers, Systems Software | 15-1133 | Software Engineer | 85610 | 110000 | 139550 |
| IT Security Specialist | Security Management Specialists | 13-1199 | NA | 52200 | 70530 | 94890 |
| IT Systems Engineer | Computer and Information Systems Managers | 11-3021 | NA | 110110 | 142530 | 180190 |
| IT Systems Engineer | Information Technology Project Managers | 15-1199 | IT Specialist | 66410 | 90270 | 117070 |
| IT Project Manager | Computer and Information Systems Managers | 11-3021 | NA | 110110 | 142530 | 180190 |
| IT Project Manager | Information Technology Project Managers | 15-1199 | IS/IT Project Manager | 66410 | 90270 | 117070 |
| Project Manager | Construction Managers | 11-9021 | Project Manager | 70670 | 93370 | 123720 |
| Project Manager | Architectural and Engineering Managers | 11-9041 | Project Manager | 112400 | 140760 | 173180 |
| Project Manager | Managers, All Other | 11-9199 | Project Manager | 75460 | 107480 | 143230 |
| Project Manager | Information Technology Project Managers | 15-1199 | Project Manager | 66410 | 90270 | 117070 |
| Project Manager | Environmental Engineers | 17-2081 | Project Manager | 66590 | 87620 | 112230 |
| Project Manager | Wind Energy Engineers | 17-2199 | Project Manager | 69890 | 96980 | 126200 |
| Project Manager | Environmental Restoration Planners | 19-2041 | Project Manager | 53580 | 71130 | 94590 |
| Project Manager | Social Science Research Assistants | 19-4061 | Project Manager | 35450 | 46640 | 60830 |
| Project Manager | Remote Sensing Technicians | 19-4099 | Project Manager | 37940 | 49670 | 63340 |
| Project Manager | Technical Directors/Managers | 27-2012 | Project Manager | 48520 | 71680 | 110350 |
| Project Manager | Intelligence Analysts | 33-3021 | Project Manager | 57560 | 81920 | 107000 |
| Senior Staff | NA | NA | NA | NA | NA | NA |
| Policy Specialist | NA | NA | NA | NA | NA | NA |
| Administrative Staff | First-Line Supervisors of Office and Administrative Support Workers | 43-1011 | NA | 42750 | 55810 | 71550 |
| Administrative Staff | Executive Secretaries and Executive Administrative Assistants | 43-6011 | NA | 46530 | 59340 | 74460 |
| Administrative Staff | Secretaries and Administrative Assistants, Except Legal, Medical, and Executive | 43-6014 | NA | 28930 | 36630 | 46230 |
| Administrative Staff | Business Operations Specialists, All Other | 13-1199 | Administrative Assistant | 52200 | 70530 | 94890 |
| Administrative Staff | Billing and Posting Clerks | 43-3021 | Administrative Assistant | 31870 | 37800 | 46350 |
| Administrative Staff | New Accounts Clerks | 43-4141 | Administrative Assistant | 30300 | 35800 | 42050 |
| Administrative Staff | Medical Secretaries | 43-6013 | Administrative Assistant | 29580 | 35760 | 43200 |
| Facilities Manager | General and Operations Managers | 11-1021 | Facilities Manager | 65650 | 100930 | 157120 |
| Facilities Manager | Administrative Services Managers | 11-3011 | Facilities Manager | 71850 | 96180 | 127100 |
| Facilities Manager | Property, Real Estate, and Community Association Managers | 11-9141 | Facilities Manager | 41210 | 58340 | 85120 |
| Facilities Manager | First-Line Supervisors of Housekeeping and Janitorial Workers | 37-1011 | Facilities Manager | 31020 | 39940 | 52280 |
| Facilities Manager | First-Line Supervisors of Office and Administrative Support Workers | 43-1011 | Facilities Manager | 42750 | 55810 | 71550 |
| Facilities Manager | First-Line Supervisors of Mechanics, Installers, and Repairers | 49-1011 | Facilities Manager | 51430 | 66140 | 83980 |
| Facilities Manager | Maintenance and Repair Workers, General | 49-9071 | Facilities Manager | 29560 | 38300 | 50100 |
| Data Scientist | Computer and Information Research Scientists | 15-1111 | Data Scientist | 91650 | 118370 | 149470 |

Table 2 ([nlm.collapsed.csv](../analysis/nlm.collapsed.csv))lists the ranges, as defined above, for each of the normative description, based on the underlying occupations identified.

|  |  |  |  |
| --- | --- | --- | --- |
| Job Title | PCT25 | MEDIAN | PCT75 |
| Administrative Staff | 28930 | 37800 | 94890 |
| Communication Specialist | 44490 | 60000 | 81550 |
| Curator | 38090 | 53780 | 80230 |
| Data Librarian | 38090 | 59050 | 90550 |
| Data Scientist | 91650 | 118370 | 149470 |
| Data Wrangler | 66410 | 90270 | 117070 |
| Education Specialist | 39800 | 60600 | 82860 |
| Facilities Manager | 29560 | 58340 | 157120 |
| Informatician | 66410 | 89505 | 117070 |
| IT Project Manager | 66410 | 116400 | 180190 |
| IT Security Specialist | 52200 | 70530 | 94890 |
| IT Systems Engineer | 66410 | 116400 | 180190 |
| Metadata Librarian | 38090 | 59050 | 90550 |
| Policy Specialist | Inf | NA | -Inf |
| Project Manager | 35450 | 87620 | 173180 |
| Records Management Specialist | 38090 | 59050 | 90550 |
| Research Domain Curator | 47500 | 80300 | 173180 |
| Research Domain Project Manager | 47500 | 80300 | 173180 |
| Researcher | 40670 | 79945 | 173180 |
| Senior Staff | Inf | NA | -Inf |
| Software Engineer | 79340 | 110000 | 149470 |

Table 3 ([nlm.categories.csv](../analysis/nlm.categories.csv))lists the statistics associated with each of the L-VH categories. While we defined the categories based on our own experience, ex ante, they match up well with observed median salaries in 2018.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Relative Salary | PCT25 | MEDIAN | PCT75 | N | Missing |
| L | 28930 | 37800 | 94890 | 7 | 0 |
| M | 29560 | 61505 | 173180 | 34 | 0 |
| H | 40670 | 80300 | 180190 | 50 | 1 |
| VH | 52200 | 103620 | 180190 | 10 | 1 |

# Full code and data

The code and data underlying this chapter, including an exhaustive list of our own edits (inclusions and exclusions) to the list of the occupations, is available at <https://github.com/labordynamicsinstitute/job-description-and-wages> and (DOI TBD).

# References

Bureau of Labor Statistics. 2019a. “2018 Standard Occupational Classification System.” <https://www.bls.gov/soc/2018/major_groups.htm>.

———. 2019b. “Occupational Employment Statistics.” Dataset. Bureau of Labor Statistics, OES Program. <https://www.bls.gov/oes/home.htm>.

———. 2019c. “Survey Methods and Reliability Statement for the May 2018 Occupational Employment Statistics Survey.” Bureau of Labor Statistics, OES Program. <https://www.bls.gov/oes/current/methods_statement.pdf>.

National Center for O\*NET Development. 2019a. “O\*NET Data Collection Overview.” <https://www.onetcenter.org/dataCollection.html>.

———. 2019b. “O\*NET® 23.2 Database.” Data Dictionary. O\*NET Resource Center. <https://www.onetcenter.org/dl_files/database/db_23_2_dictionary.pdf>.

———. 2019c. “O\*NET® Database Release 23.2.” Dataset. O\*NET Resource Center. <https://www.onetcenter.org/db_releases.html>.

Robinson, David. 2019. *Fuzzyjoin: Join Tables Together on Inexact Matching*. <https://github.com/dgrtwo/fuzzyjoin>.

Winkler, William E. 1990. “String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage.” *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 354–59. <https://files.eric.ed.gov/fulltext/ED325505.pdf>.