Dictionary of Occupational Titles (DOT), 3rd Edition

César Garro-Marín[*], Vittoria Dicandia[†], Costas Cavounidis[‡], Kevin Lang[§], and Raghav Malhotra[¶]

December 23, 2021

1 Overview

This repository contains digitized data on job skill requirements coming from the third Edition of the *Dictionary of Occupational Titles (DOT)*. This dataset contains information on jobs and its skill requirements. We structured it to match as closely as possible the publicly available data from the *Fourth Edition* (National Academy of Sciences, 1971). Whenever possible, we matched variable coding and names with those from the fourth edition.

2 What does the DOT contain?

The DOT contains information on jobs and functions that workers perform in these jobs. The DOT identifies each job with a six digit number. For example, the job *personal shopper* corresponds to the code 296.358. The first three digits indicate the occupational group the job belongs to. In this example 296 corresponds to the group *Shoppers*. The last digits identify **functions** or **tasks** that workers must perform in this job. Here, 358 identifies the function *demonstration and sales work*. Note that a function can be shared by several job titles. For example, besides personal shoppers, *demonstrations and sales work* is also performed by *Bridal consultants* (299.358).

For each *function*, the DOT provides descriptions of the work performed, worker requirements, and numerical scales describing the qualification profile required for an average performance in the job. Figure 1 shows an example. The qualification profiles contain information on:

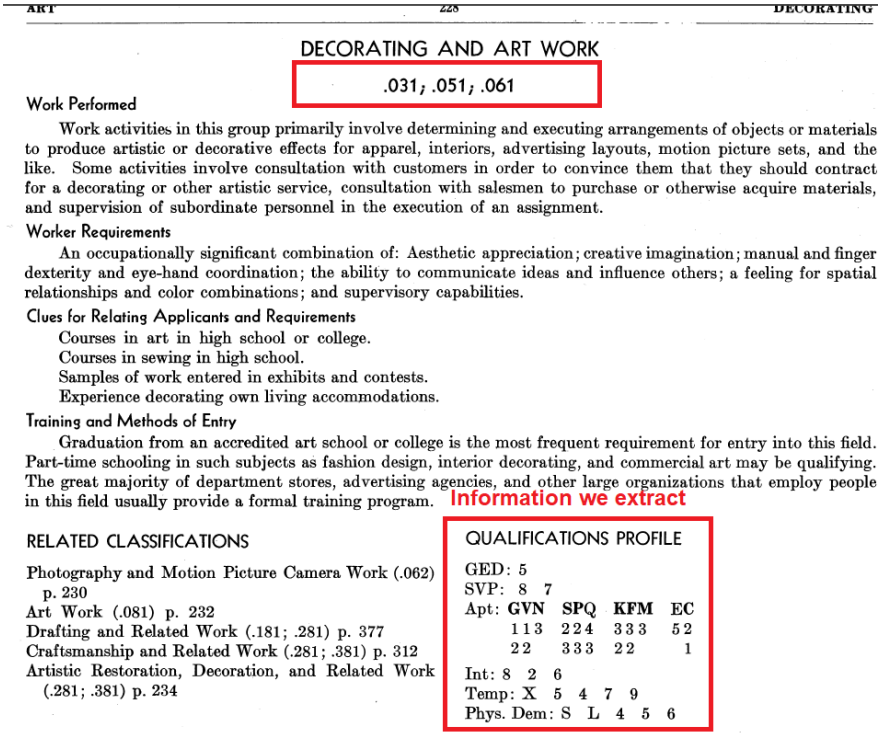
- **Educational requirements:** general education development (GED) and specific vocational preparation (SVP).
- **Aptitudes:** intelligence (G), verbal (V), numerical (N), spatial (S), form perception (P), clerical perception (Q), motor coordination (K), finger dexterity (F), manual dexterity (M), eye-hand-foot coordination (E), color discrimination (C).

[*]Boston University, email: cesarlgm@bu.edu
[†]Boston University, email: dicandia@bu.edu
[‡]Warwick University, email: C.Cavounidis@warwick.ac.uk
[§]Boston University, email: lang@bu.edu
[¶]Warwick University, email: r.malhotra@warwick.ac.uk

- **Interests.**
- **Temperaments.**
- **Physical demands.**

This qualification profiles are the information we digitized and made available in the accompanying datasets.

Figure 1: Qualification profile example



ART 228 DECORATING

DECORATING AND ART WORK

.031; .051; .061

Work Performed

Work activities in this group primarily involve determining and executing arrangements of objects or materials to produce artistic or decorative effects for apparel, interiors, advertising layouts, motion picture sets, and the like. Some activities involve consultation with customers in order to convince them that they should contract for a decorating or other artistic service, consultation with salesmen to purchase or otherwise acquire materials, and supervision of subordinate personnel in the execution of an assignment.

Worker Requirements

An occupationally significant combination of: Aesthetic appreciation; creative imagination; manual and finger dexterity and eye-hand coordination; the ability to communicate ideas and influence others; a feeling for spatial relationships and color combinations; and supervisory capabilities.

Clues for Relating Applicants and Requirements

Courses in art in high school or college.
Courses in sewing in high school.
Samples of work entered in exhibits and contests.
Experience decorating own living accommodations.

Training and Methods of Entry

Graduation from an accredited art school or college is the most frequent requirement for entry into this field. Part-time schooling in such subjects as fashion design, interior decorating, and commercial art may be qualifying. The great majority of department stores, advertising agencies, and other large organizations that employ people in this field usually provide a formal training program.

Information we extract

RELATED CLASSIFICATIONS

Photography and Motion Picture Camera Work (.062) p. 230
Art Work (.081) p. 232
Drafting and Related Work (.181; .281) p. 377
Craftsmanship and Related Work (.281; .381) p. 312
Artistic Restoration, Decoration, and Related Work (.281; .381) p. 234

QUALIFICATIONS PROFILE

GED: 5
SVP: 8 7
Apt: **GVN SPQ KFM EC**
1 1 3 2 2 4 3 3 3 5 2
2 2 3 3 3 2 2 1
Int: 8 2 6
Temp: X 5 4 7 9
Phys. Dem: S L 4 5 6

3 How did we structure the digitized information?

To accommodate the needs of different users, we released two sets of files. Ready to use files provide all the DOT information along with cross-walks to link them to the census or CPS files. Users looking for the raw data as it comes from the DOT should refer to the long format files.

Table 1: Release description: DOT 3rd edition

File name	Contents	Observations
Folder `code/`		
reshape_mmm_dot.do	Do file that creates dot1965_summary_v1.dta and dot1965_wide_v1.dta using the long format files	NA
Folder `data/ready_to_use_files/`		
cw_dot65_census1970.dta	Cross-walk between 1970 census and DOT 1965 occupation codes	7,095
dot1965_summary_v1.dta	Dataset with all DOT information. It contains average, max and min qualification requirements by job title*	8,147
dot1965_wide_v1.dta	Dataset with all DOT information. It contains all qualification requirements by job title in wide format*	8,147
Folder `data/long_format_files/`		
worker_function_long_v1.dta	Job's relationship to data, people and things	3,234
ged_long_v1.dta	General educational development	6,167
svp_long_v1.dta	Specific vocational preparation	9,888
aptitudes_long_v1.dta	Aptitudes	8,281
interests_long_v1.dta	Interests	3,251
temp_long_v1.dta	Temperaments	3,234
physdem_long_v1.dta	Physical demands	10,564

Notes: these datasets contain all the information we digitized from the DOT, 3rd edition.

* A job title can contain several lines of data for the same requirement. This is because a task can be performed at several "several levels of complexity". Whenever this happens, the file `dot1965_wide_v1.dta` contains all job-title requirements in a wide format. The file `dot1965_summary_v1` contains averages, maximum and minimums of these requirements by `job_title` and `task_id`.

3.1 Ready to use files

The folder `data/ready_to_use_files/` contains three datasets:

- `cw_dot65_census1970.dta`: contains a cross-walk between 1970 Census occupation codes and the 1965 DOT occupation codes (i.e the three first digits of the job title). An observation in this dataset is a census occupation-DOT job title pair.

  We built the crosswalk using the augmented 1971 CPS (National Academy of Sciences, 1971).

This CPS contains both the DOT job titles and Census occupation codes, along with information on employment. Typically, a DOT job title maps into several census occupations. In these cases, we weight each Census-DOT code pair using the employment share they occupy within the census occupation. That is,

$$cw\_weight_{ij} = \frac{N_{ij}}{\sum_l N_{il}}$$

where N denotes employment, i is the *census occupation* and j is the *DOT job title*.

- `dot_1965_summary_v1.dta`: contains summary information from the job qualification profiles. The file also includes the DOT-census cross-walk, thus and observation is indexed by the pair of census occupation-DOT job title.

  Some job tasks can be performed at different levels of complexity. Therefore, a job task could feature multiple lines for the same requirement. For example, figure 1 shows that *Decorating and Artwork* can require either a minimum of 2 to 4 years (SVP 7) or 4 to 10 years (SVP 8) of specific vocational preparation. In this dataset we provide several summaries of these multiple levels. We computed averages, maximum and minimums for each possible requirement. These summary variables are identified with the appropriate suffix. Variables without any suffix could have only one value per job title. See section 3.2 for a detailed description of what each requirement measures.

- `dot_1965_summary_v1.dta`: contains job qualification profile requirements from 1965 DOT in a wide format. The file also includes the DOT-census cross-walk, thus and observation in this dataset is indexed by the pair census occupation-DOT job title.

  This file contains all the data by job title in a wide format. Therefore, whenever a given skill requirement had different levels in the data, all the possible levels appear in this dataset indexed by a suffix. For example, if for a job title `ged` (General Education Development) had two levels, they will appear in the variables `ged1` and `ged2`. Variables without any suffix had only one value per job title. See section 3.2 for a detailed description of what each requirement measures.

3.2 Long format files

The long format files divide the DOT data into seven files. Table 1 provides a brief description of the contents from each file. The user should use variables `job_title` and `task_id` to link all the datasets. In the next subsections, we describe each file in more detail.

As mentioned before, a task can feature several levels of the same requirement. In these files we provide all requirement levels in a long format. Requirement levels are indexed by the variable `requirement`.

3.3 Relation to data, people and things: worker_function_long_v1.dta

This dataset contains information on how the job relates to data, people and things. The last three digits of the job title code express the job's relationship to Data, People and Things in that order. These digits express the most complex function that the job requires from the worker. The combination of `job_title` and `task_id` identifies observations in this dataset. Table 2 provides a brief description of the files' contents. Table 3 shows the codebook for the relevant variables.

Table 2: Variable description: worker_function_long_v1.dta

Variable name	**Description**	**Type**	**Range**
job_title	Job title	string	
task_id	Task identifier	integer	
occ_group	Occupational group	string	
worker_function	Worker function	string	
data	Relation to data	integer	0-8
people	Relation to people	integer	0-8
things	Relation to things	integer	0-8

Note: the combination of `job title` and `task_id` indexes observations in this dataset.

Table 3: Codebook: worker_function_long_v1.dta

Variable: data		*Variable: people*		*Variable: things*	
0	Synthesizing	0	Mentoring	0	Setting-up
1	Coordinating	1	Negotiating	1	Precision working
2	Analyzing	2	Instructing	2	Operating-controlling
3	Compiling	3	Supervising	3	Driving-operating
4	Computing	4	Diverting	4	Manipulating
5	Copying	5	Persuading	5	Tending
6	Comparing	6	Speaking-signaling	6	Feeding-off bearing
7	No relationship	7	Serving	7	Handling
8	No relationship	8	No relationship	8	No relationship

3.4 General Education Development: ged_long_v1.dta

This dataset contains information on the general educational requirements for average performance on the job. An observation in this dataset is indexed by the combination of `job_title`, `task_id` and `requirement`. Table 4 provides a brief description the variables in this file.

Table 4: Variable description: ged_long_v1.dta

Variable name	Description	Type	Range
job_title	Job title	string	
task_id	Task identifier	integer	
occ_group	Occupational group	string	
worker_function	Worker function	string	
requirement	Requirement level	integer	1-3
ged	Amount of general education the worker requires for average performance in the job.	integer	1-6

Note: the combination of `job title`, `task_id`, and `requirement` indexes observations in this dataset.

The values of the `ged` variable describe education requirements on three dimensions: reasoning, mathematical, and language development. Figure 2 provides a detailed description of what each GED code entails.

Figure 2: Description of General Educational Development levels

GENERAL EDUCATIONAL DEVELOPMENT

Level	Reasoning Development	Mathematical Development	Language Development
6	Apply principles of logical or scientific thinking to a wide range of intellectual and practical problems. Deal with nonverbal symbolism (formulas, scientific equations, graphs, musical notes, etc.) in its most difficult phases. Deal with a variety of abstract and concrete variables. Apprehend the most abstruse classes of concepts.	Apply knowledge of advanced mathematical and statistical techniques such as differential and integral calculus, factor analysis, and probability determination, or work with a wide variety of theoretical mathematical concepts and make original applications of mathematical procedures, as in empirical and differential equations.	Comprehension and expression of a level to —Report, write, or edit articles for such publications as newspapers, magazines, and technical or scientific journals. Prepare and draw up deeds, leases, wills, mortgages, and contracts. —Prepare and deliver lectures on politics, economics, education, or science. —Interview, counsel, or advise such people as students, clients, or patients, in such matters as welfare eligibility, vocational rehabilitation, mental hygeine, or marital relations. —Evaluate engineering technical data to design buildings and bridges.
5	Apply principles of logical or scientific thinking to define problems, collect data, establish facts, and draw valid conclusions. Interpret an extensive variety of technical instructions, in books, manuals, and mathematical or diagrammatic form. Deal with several abstract and concrete variables.		
4	Apply principles of rational systems [1] to solve practical problems and deal with a variety of concrete variables in situations where only limited standardization exists. Interpret a variety of instructions furnished in written, oral, diagrammatic, or schedule form.	Perform ordinary arithmetic, algebraic, and geometric procedures in standard, practical applications.	Comprehension and expression of a level to —Transcribe dictation, make appointments for executive and handle his personal mail, interview and screen people wishing to speak to him, and write routine correspondence on own initiative. —Interview job applicants to determine work best suited for their abilities and experience, and contact employers to interest them in services of agency. —Interpret technical manuals as well as drawings and specifications, such as layouts, blueprints, and schematics.
3	Apply common sense understanding to carry out instructions furnished in written, oral, or diagrammatic form. Deal with problems involving several concrete variables in or from standardized situations.	Make arithmetic calculations involving fractions, decimals and percentages.	Comprehension and expression of a level to —File, post, and mail such material as forms, checks, receipts, and bills. —Copy data from one record to another, fill in report forms, and type all work from rough draft or corrected copy. —Interview members of household to obtain such information as age, occupation, and number of children, to be used as data for surveys, or economic studies. —Guide people on tours through historical or public buildings, describing such features as size, value, and points of interest.
2	Apply common sense understanding to carry out detailed but uninvolved written or oral instructions. Deal with problems involving a few concrete variables in or from standardized situations.	Use arithmetic to add, subtract, multiply, and divide whole numbers.	
1	Apply common sense understanding to carry out simple one- or two-step instructions. Deal with standardized situations with occasional or no variables in or from these situations encountered on the job.	Perform simple addition and subtraction, reading and copying of figures, or counting and recording.	Comprehension and expression of a level to —Learn job duties from oral instructions or demonstration. —Write identifying information, such as name and address of customer. weight, number, or type of product, on tags, or slips. —Request orally, or in writing, such supplies as linen, soap, or work materials.

[1] Examples of "principles of rational systems" are: Bookkeeping, internal combustion engines, electric wiring systems, house building, nursing, farm management, ship sailing.

Source: US Department of Labor (1965).

3.5 Specific Vocational Preparation: svp_long_v1.dta

This dataset contains information on the amount of time required to learn the techniques, acquire information and develop the facility needed for average performance on the job. An observation in this dataset is given by the combination of `job_title`, `task_id` and `requirement`. Table 5 provides a brief description of all the variables in this dataset. Table 6 shows the codebook for the relevant variables.

Table 5: Variable description: svp_long_v1.dta

Variable name	Description	Type	Range
job_title	Job title	string	
task_id	Task identifier	integer	
requirement	Requirement level	integer	1-5
occ_group	Occupational group	string	
worker_function	Worker function	string	
svp	Amount of time of specific vocational preparation required for the job	integer	1-9

Note: the combination of `job title`, `task_id`, and `requirement` indexes observations in this dataset.

Table 6: Codebook: svp_long_v1.dta

Code	Description
Variable: svp	
1	Short demonstrations only
2	Anything beyond short demonstrations up and including 30 days.
3	Over 30 days up to 3 months.
4	Over 3 months up to 6 months.
5	Over 6 months up to 1 year.
6	Over 1 year and up to 2 years
7	Over 2 years up to 4 years.
8	Over 4 years up to 10 years.
9	Over 10 years.

3.6 Aptitudes: aptitudes_long_v1.dta

This dataset contains information describing specific capacities and abilities required to perform adequately the job. An observation in this dataset is given by the combination of `job_title`, `task_id`, and `requirement`. Table 7 provides a brief description of all the variables in this dataset. Table 8 shows the codebook for the relevant variables.

Table 7: Variable description: aptitudes_long_v1.dta

Variable	Description	Type	Range
job_title	Job title	string	
task_id	Task identifier	integer	
requirement	Requirement level	integer	1-4
occ_group	Occupational group	string	
worker_function	Worker function	string	
intelligence	General ability to "catch on" or understand instructions and underlying principles.	Integer	1-4
verbal	Ability to understand meanings of words and ideas associated with them	Integer	1-4
numerical	Ability to perform arithmetic operations quickly and accurately	Integer	1-5
spatial	Ability to comprehend forms in space and understand relationships of plane and solid objects	Integer	1-5
form_perception	Ability to perceive detail in objects of in pictorial or graphic material	Integer	1-5
clerical_perception	Ability to perceive pertinent detail in verbal or tabular material	Integer	1-5
motor_coordination	Ability to coordinate eyes and hands or fingers rapidly and accurately in making precise movements with speed	Integer	1-5
finger_dexterity	Ability to move the fingers and manipulate small objects with the fingers rapidly or accurately.	Integer	1-5
manual_dexterity	Ability to move the hands easily and skillfully.	Integer	1-5
eye_hand_foot coordination	Ability to move the hand and foot coordinately with each other in accordance with visual stimuli	Integer	1-5
color_discr	Ability to perceive or recognize similarities or differences in colors.	Integer	1-5
intelligence_f	Is intelligence essential for performing the job	Integer	0-1
verbal_f	Is verbal aptitude essential for performing the job	Integer	0-1
numerical_f	Is numerical aptitude essential for performing the job	Integer	0-1
spatial_f	Is spatial aptitude essential for performing the job	Integer	0-1
form_perception_f	Is form perception essential for performing the job	Integer	0-1
clerical_perception_f	Is clerical perception essential for performing the job	Integer	0-1
motor_coordination_f	Is motor coordination essential for performing the job	Integer	0-1
finger_dexterity_f	Is finger dexterity essential for performing the job	Integer	0-1
manual_dexterity_f	Is manual dexterity essential for performing the job	Integer	0-1
eye_hand_foot_f	Is eye-hand-foot coordination essential for performing the job	Integer	0-1

Note: the combination of `job title`, `task_id`, and `requirement` indexes observations in this dataset.

Table 7: Variable description: aptitudes_long_v1.dta

Variable	Description	Type	Range
color_discr_f	Is color discrimination essential for performing the job	Integer	0-1

Note: the combination of `job title`, `task_id`, and `requirement` indexes observations in this dataset.

Table 8: Codebook: aptitudes_long_v1.dta

Code	Description
Variables: intelligence-color_discr	
1	Requirements at the top 10 percent of the population.
2	Requirements at top third of the population, exclusive to the top 10 percent.
3	Middle third of the population.
4	Lowest third of the population, exclusive of the bottom 10 percent
5	Lowest 10 percent of the population.
Variables: intelligence_f-color_discr_f	
0	Non-essential
1	Essential

3.7 Interests: interests_long_v1.dta

This dataset contains information on preferences for certain types of work activities or experiences. An observation in this dataset is given by the combination of `job_title`, `task_id`, and `requirement`. Table 9 provides a brief description of all the variables in this dataset.

The five variables from `p_people` to `p_tangible` indicate preference for a certain kind of activity. Each of these variables involve the comparison of a pair of activities. Declaring preference for one activity implies rejection of the other activity in the pair. For example, in `p_people`, declaring a preference for "activities concerned with people and the communication of ideas" implies a rejection of activities "dealing with things and objects". The codebook in table 10 describes the situations involved in each pair.

Table 9: Variable description: interests_long_v1.dta

Variable	Description	Type	Range
job_title	Job title	string	
task_id	Task identifier	integer	
occ_group	Occupational group	string	
worker_function	Worker function	string	
p_people	Things and objects **vs** people	Int	-1-1
p_science	Business contact with people **vs** scientific activities	Int	-1-1
p_abstract	Routine activities **vs** abstract activities	Int	-1-1
p_non_social	Working for the social good **vs** non-social activities	Int	-1-1
p_tangible	Prestige **vs** tangible, productive satisfaction	Int	-1-1

Note: the combination of `job title` and `task_id` indexes observations in this dataset.

Table 10: Codebook: interests_long_v1.dta

Code	Description
Variable: p_people	
-1	Preference for activities dealing with things and objects
0	Neutral
1	Preference for activities concerned with people and the communication of ideas
Variable: p_science	
-1	Preference for activities involving business contact with people
0	Neutral
1	Preference for activities of a scientific and technical nature
Variable: p_abstract	
-1	Preference for activities of a routine, concrete, organized nature
0	Neutral
1	Preference for activities of abstract and creative nature
Variable: p_non_social	
-1	Preference for working for people, for their presumed good, as in the social welfare sense, or for dealing with people and language
0	Neutral
1	Preference for activities that non-social in nature, and are carried on in relation to processes, machines and techniques
Variable: p_tangible	
-1	Situations involving a preference for activities resulting in prestige or the esteem of others
0	Neutral
1	Situations involving a preference for activities resulting in tangible, productive satisfaction

3.8 Temperaments: temp_long_v1.dta

This dataset contains information on the presence of certain types of occupational situations to which workers must adapt. The combination of `job_title` and `task_id` indicates and observation in this dataset. Table 11 provides a brief description of all the variables in the file. All the variables from `varch` to `sts` are dummies indicating the presence of the occupational situations from the description.

Table 11: Variable description: temp_long_v1.dta

Variable	Description	Type	Range
job_title	Job title	string	
task_id	Task identifier	integer	
occ_group	Occupational group	string	
worker_function	Worker function	string	
varch	Variety of duties often characterized by frequent change	Int	0-1
repcon	Repetitive or short cycle operations carried out according to set procedures or sequences	Int	0-1
no_discretion	Doing things only under specific instruction, allowing little or no room for independent action or judgment in working out job problems	Int	0-1
dcp	Direction, control and planning of an entire activity or the activities or others	Int	0-1
dpl	Necessity of dealing with people in actual job duties beyond giving and receiving instructions	Int	0-1
alone	Working alone and apart in physical isolation from others, although the activity may be integrated with that of others	Int	0-1
influ	Influencing people in their opinions, attitudes, or judgments about ideas or things	Int	0-1
pus	Performing adequately under stress when confronted with the critical or unexpected or when taking risks	Int	0-1
sjc	Evaluation (arriving at generalizations, judgments, or decisions) of information against sensory or judgmental criteria	Int	0-1
mvc	Evaluation (arriving at generalizations, judgments, or decisions) of information against measurable and verifiable criteria	Int	0-1
fif	Interpretation of feelings, ideas, or facts in terms of personal viewpoint	Int	0-1
sts	Precise attainment of set limits, tolerances, or standards	Int	0-1

Note: the combination of `job title` and `task_id` indexes observations in this dataset.

3.9 Temperaments: physdem_long_v1.dta

This dataset contains information on the physical activities required of a worker in a job. The combination of `job_title` and `task_id` indicates and observation in this dataset. Table 12 provides a brief description of its variables. All the variables from `climb` to `see` are dummies indicating that the physical activity in the description is required. Moreover, table 13 shows the codebook for `strength` scale.

Table 12: Variable description: physdem_long_v1.dta

Variable	Description	Type	Range
job_title	Job title	string	
task_id	Job title	integer	
occ_group	Occupational group	string	
worker_function	Worker function	string	
strength	Physical strength requirements	Int	1-5
climb	Climbing or balancing	Int	1-2
stoop	Stooping, kneeling, crouching or crawling	Int	1-4
reach	Reaching, handling, fingering, or feeling	Int	1-4
talk	Talking or hearing	Int	1-2
see	Seeing	Int	1-5

Note: the combination of `job title` and `task_id` indexes observations in this dataset.

Table 13: Codebook: physdem_long_v1.dta

Code	Description
Variable: strength	
1	Sedentary work
2	Light work
3	Medium work
4	Heavy work
5	Very heavy work

4 Digitization process

We digitized the database in three steps. First, we scanned all the job function pages available in the DOT. These pages contain textual descriptions tasks and worker requirements, along with a table containing the *qualification profile* (see figure 1 for an example). Next, we extracted all the information from the qualification profile table using the OCR engine tesseract. Finally, to guarantee data quality, we checked all the extracted data manually. We corrected the mistakes from the OCR engine manually. All the code and documentation from the scanning process is available at https://github.com/cesarlgm/dot_1965_creation.

References

National Academy of Sciences. Committee on Occupational Classification and Analysis. (1971). *Dictionary of Occupational Titles (DOT): Part I - Current Population Survey , April 1971 , Augmented With DOT Characteristics and Dictionary of Occupational Titles (DOT): Part II - Fourth Edition Dictionary of DOT Scores for 1970.* Number April. Inter-university Consortium for Political and Social Research [distributor].

US Department of Labor (1965). *Dictionary of Occupational Titles.* U.S. Government Printing Office, Washington, D.C., 3rd edition.