

# Practicum 3: Big Data

In this week's practicum we will be asking Python to process administrative data for us. Specifically, we will count how many persons of different demographic groups live in our little town and compute Labor Force Participation Rates. We will be working with a manageable-sized subset of data, but keep an eye out for pieces of code that would be extensible to larger datasets.

## 1 Demographics and Tax Collection in Pahrump, Nevada 1920

Data collection for public administration is, one could say, a progenitor of Data Science. Many of the basic tasks of public administration, such as counting people and calculating taxes, involve collecting and summarizing data. Collecting accurate administrative data is an important component of conducting accurate analyses, which allow the powers that be to make informed decisions. Censuses are a specific kind of administrative data collection, where every effort is made to collect the same basic information from every resident of a polity. Censuses have been conducted since ancient times, notably in Babylon, Persia, and Rome.

The US Census is conducted every 10 years and is used in many aspects of public administration. By law, US Census individual records (ie. 'microdata') are kept behind lock and key for 72 years. After this they become available for demographic, genealogical, and other study. We will use 1920 US Census data to characterize the population of Pahrump, Nevada at that point in time.

TOWNSHIP	NAME OF INST	-	Ely lity				manusal	NAME OF INCOPPOR	ATED PLACE	CN THE 28	-	сь Даниали	1000 TK			1	722	4
PLACE OF	ARCON.	700	BHATIST.	DRK.	TORK COPTA	CITIZENSE	P. BECATIES.			NATIONAL AND MA	OTHER PERSON.	0 1	- 0 - 0	Ti		exercises.	- L	7
T	1 mm	person whose place of abods on 17 L 1988, was in this family.		1. 14	1 1 1	M. 1 1	114 1 1	See Children you				orientes, Dallangalista,		14			ber 6	٥
I E	1	AND THE PERSON NAMED IN	- Barrier etts	in in	1 10 10		1	Passent.	-	Park No.	Beharings.	Paradition.	Televines.	۱ï	,Biance	THE S	· == "	1
			- 1,		* D H 1	0 11	5 18 17 18	19			30	10		-		11		
	-	- Gleritiyi	Wiefer Outstan		F w 29 M	+++	74.70	Manuelas	-	New York	-	needs	-	-	Moute.	-	+	+
		- Jenseyie	Ql. elk.	ш	PW 7 a		Soul !	Herada.		Menada.		Musla		0	Henry			_;
	www Porns	William	Tient 1	0 1	M W SE M	4	200	Hevefa	-	Gurgder .	· English	Menda Menda England	Suchil		Bertheren	1010	-	1:
m 3	7		76.6		F W 32 M	1 84	Och	Halistot		Annany	Home	Layered	Grower	See.	Moule	Municipal		₫; '
1	-	- Though	8 14	Н	F W 7 0	-	Jan of	With,	-	Test Jerdey	-	Kauses	-	4	House.	-	-	+
NX s	N SHE Oak	1. Vote Bulgie	Stylle,	ę	M W W M	. 1 1	100 1500	Tellenia				Тиши		4	Popute Ship	1xlATing	dw.	_;
( )	-	Better	The fitter	Н	6 W 25 M	4	gulya	Durada	-	Missourie	-	Colorado	-	gu	Mark.		1	-10
1	Rolet	Muckey	Vietelin lear		F w 15 10		No 100 100	The Moure	-	Dhasami			-		Your		世	<b>_</b> ';
140	27 359 Beile	Marie	Hiller 1	8.	M W 25 M	4	4 Spelyer	Ohis .	-	Phio	-	Treback	2 de	14	Samples.	Reduction	1/4	Ë
3		- 16 than D.	dele		Mw Istil a					Bungaitt				1	There			1,5
12	100	Les Wargenet	Soul.	ш	Mw Si.	2000 000 0		Burglas Greland Cugland Sudiana	Brief ,	Consequity		Telayd England	0 /			-		
1	21 50 Can	Manie 102	Heat 1	b #	K W 50 W	o la art Am la	o della	Quelent	Queglish	England	Bushish	England	Qualit	1	Buckey	Quality	ch	17
X.	28 257 Noche		There !	8	Mw W M		Spectice	ofediane.		Hie	1	Indiana.	/	gu.	Thinte	Nouse!	PM.	29 -
		- Ange	The.	ш	M w 2/ 10					Gedgain.		Queliaga.		1	Secretar.	House	w	20
	-	- Clayerer	ston,		19 W 1/9 W		Sperger	Queliana.		Queliaina		Gerdinia.	-	344	Beeled Builton Beauten Builton Builton Beauten Beauten Beauten	House	w	22
1		- Last	OLLA		F W 17 e		Androdon.	P. Leure		Sections.		Queliana.		100	House		+	22
		- Depuly	afre	$\perp$	M W 11 4	4	Andready	Blake		Dediana	-	Quelegia		Jan.	Moure.			28
		Burbery	Qualte	Н	EW VI		119	Musda		Zydiana.		Gederica		1	Your Yours Venes		+	20
	70	- 16 Rus	Block		M W 39 E	4	yeary	Quediaire.		Geologica. Geologica	0 11	Judiaire		See	Menter Done	Lower	W.	- 28
1	20 252 Calor	- (Kickens)	Topal 1	R.	F W 36 W	0	San Market	Musela.		Sperieni	Rylich	Neveda	English	d'in	There		+	20
N	1	- (Kobert	Dones		mw H &		Buchalan	Colorado		Muconein	0 11	Murda	- 11	Jan.	Henel ,			n
k 2	See Steel	w Fraker	Brother.	2 5	F W 66 W			Blonedo .	-	Grefand	Sugare	Gueland	English	200	Auto Mechan	garage	W/	12
43	-	- Haveld	Same		MW 29 4		Spelype	Michani			0.77			de	There,			1,4
2 183 1	way Home	. Garl.	Then !	R.	M w 25 M	2018	7000	Michael Gellenk Hellend Hele	Dutely	Halland Valland	Sutter.	12 Sand	Dutal	764	Theret.	Bekery	w	32
1		Jane .	Rose	ш	MW 50		111	Ittaly		Halland	Dated.	and and	Cutel,	4	Henry Behavi Henry Doubs			17
	12 sec 2	de William	Theel, 1	0	Mw 28 0	4	100.00	Heelington.		Typlland	Dutela	Tolland	Dutet			Reduction 20	d	- 3×
1	The second	- Genette	Make		F W 26 M		345-	Tetati.		Manuel the	+	England	Euglish		Marcel		1	- 60
3	-	- William	Solv	ш	Mw Saw	4	199	Woods	-	Markington Variantes	-	Attaly	0 0	1	Joseph Contractor	-		-6
13	wi mel	ng May	Thead . 2	8	M W 33 M		la la	Leuth brekets		Sind also		Thely Ludiana		140	Quitante	Ship Lida	w	a
	ser Chan	Penney	West.	8	F W 25 M	1	1000	Leathy Gabets California		May of the	-	Egg faut &	English		Marile.	1	St.	"
	32777500	Lidentet	7/1/4		F W 25 N		2000	Thekada		Guada.	Qualit	Muse	-	No.	There I		W.	a
-	10	- George	Deutol to				100				1			4	Model Magain	11100	H	0
X	to test Jack	- More	2000		F W 07 M		du ha	Pennsylvani Bergla Cal Joine		Hickory Lock		Hydridde Georgia Redda	1	400	Mint.	State Of These	ew.	10
		- George	Parlolto	$\Box$	F W 18 4		Ser Continue	P. 1 /2.		Benefician		Media		Bec.	Houses	100000		10



#### 1.1 Download

Please download pr03.zip from the practicum website (or Blackboard), unzip it, and move the directory/folder to where you want it in your file system. Use Atom's File > "Add Project Folder..." to open this folder and view census.py.

#### 1.2 View 1920 US Census Data

First, we must load the dataset. You should see from data import data\_header, data\_as\_list\_of\_lists at the top of your file. This tells census.py to load the custom functions from data.py named data\_header and data\_as\_list\_of\_lists. These functions take as an argument the name of the file that you want to load. Please use 'nvcensus-1920-pahrump.csv'.

You should print out the dataset and confirm that it looks like census data. Use the table below to make any notes you think you might need about the dataset, such as what the options are in each column.

Year	Subunit	FName	Surname	Age	Sex	Color	Profession
1920	PAHRUMP PRECINCT	FRANK L.	CATE	44	M	W	MINER

Table 1: 1920 US Census, Pahrump, Nevada

**Looping through lists** In Python, the main use for lists is to loop through them. This means running the same snippet of code for every element in the given list, and this is useful for all sorts of things. You can print, count, subset, or even edit each entry in the same way using a very short and concise piece of code. The syntax you would use is called a "for loop" and it goes like this:

for ELEMENT in LIST:
some\_function\_of(ELEMENT)

#### 1.3 Population

We would like to know the population of Pahrump, NV in 1920 (ie. the number of records). Please print.

#### 1.4 Birth of a Child

Congratulations Pahrump! WILLIE and LUCY GOODWIN just had a daughter! The town would like to keep their records updated, so please add little ANNIE to the list and confirm that the population rose by 1.

#### 1.5 Labor Force Participation

Along with the total population, we would also like to know the Labor Force Participation Rate. To calculate this, please find the working population, which is the number of residents who have a listed Profession (ie. one that is not 'NONE'), and divide by the total population.

Please have your code print this number. To check your code, make sure this number makes sense from your understanding of the dataset.

Indexing into lists To get the Profession from the record of one of the respondents, you will need to figure out what index in the record (a list) this corresponds to. This is why it is useful to save/load/have also the header to any dataset you are using as a list. Indexing into header[i] gives you the column label for the value at record[i].

#### 1.6 Segmenting the Population

We would also like to know *who* is working. Please create four new datasets (ie. lists) for the following population subsets, and calculate the Labor Force Participation Rate for each.

Population subsets:

- Children residents under the Age of 18
- Elderly residents who are 75 years of Age or older
- Women adult residents whose Sex is F
- Men adult residents whose Sex is M

**Do these numbers make sense?** Why or why not? Please print these numbers, and write your explanation as a comment in your code.

#### 1.7 Submit Your Work!

Please submit your work to Blackboard.

If you used one file, you can upload it as-is.

If you used more, please first make a zip file of the folder within which you did your work.

- On MacOS, select your work folder, right-click and choose Compress pr03
- On Windows, select your work folder, right-click, Send To, Compressed (zipped) folder

#### 1.8 Citations

James, Ronald M.; Fliess, Kenneth H.; Nystrom, Eric C., 2014, "Nevada Census Microdata, 1860-1920", https://doi.org/10.7910/DVN/27218, Harvard Dataverse, V2

https://www.newworldencyclopedia.org/entry/Census

https://www.wikitree.com/photo/jpg/1920\_United\_States\_Federal\_Census\_Nevada\_White\_Pine\_County\_Ely\_Disrict\_0061\_p\_34

https://twitter.com/PahrumpNV

https://taxfoundation.org/us-federal-individual-income-tax-rates-history-1913-2013-nominal-and-inflation-adjusted-brackets/

## 2 Done Early?

### 2.1 Estimate of Tax Receipts

One of the age-old uses of the census is to estimate the tax base of the polity in questions. We would like to estimate the taxes that the town of Pahrump, NV could collect in 1920. See the table below for the expected contribution from each of the professions listed. Print your answer.

Profession	Salary	Tax Receipts
MINER	5000	240
FARMER	12000	1150
FARM LABORER	2600	104
OTHER PROFESSION	3200	128
NO PROFESSION	0	0

**Is your code correct?** How would you check? Would it be better to overestimate or underestimate?

### 2.2 Beyond Pahrump, NV

Within data.py there are code comments that would let you produce the corresponding dataset for any of Nevada's precincts. See if you can edit this code to do this exercise for a larger precinct.

This handout was originally created by Carolina Mattsson and Stefan Mccabe, Fall 2019.