

A blue banner with a white rounded rectangle in the top left corner containing the Willamette University logo. The logo consists of a stylized 'W' inside a circle with a compass rose design. To the right of the logo, the text 'WILLAMETTE UNIVERSITY' is written in a sans-serif font. In the center of the blue banner, the names 'Rachel Harvill, Brett Youtsey, Aaron Schechter' are written in a large, black, sans-serif font. Below the names, 'Willamette University' is written in a smaller, black, sans-serif font.

Topic

We worked collaboratively with Professor Tim Johnson from the Atkinson Graduate School of Management to analyze 40 years worth of public policy data. In particular, we worked to determine factors that affected career advancement among government employees.

We were provided with Central Personnel Data Files from 1973-2014 - which included information about all non-sensitive employees of the U.S. executive branch.

We aimed to answer our research questions while also developing a firm understanding of data science research procedures. To do so, it was necessary that we cleaned, visualized, and analyzed the data to identify patterns and relationships in the data.

Python provided the best capabilities for our analysis. We used the Pandas package to load and analyze our data. In addition, we used R to load in and convert larger data files to .csv so they could be analyzed in Python.

1. What is the influence of managers on career advancement?
2. How does an individuals entry position shape career advancement?
3. How does an individuals geographic location shape career advancement?

The original data consisted of 40 individual text files representing a year from 1973-2014. Data was coded according to Office of Personal Management standards, and not consistent throughout the years.

	0	1	2	3	4	5	6
0	000319176	009520	AA00	197309	0095	201973	0930
1	000903734	026000	AA00	197309	0260	001973	0930
2	001122701	010471	AA00	197309	0104	711973	0930
3	001585368	026000	AA00	197309	0260	001973	0930
4	001766366	020357	AA00	197309	0203	571973	0930
5	001860289	026000	AA00	197309	0260	001973	0930
6	002415114	026000	AA00	197309	0260	001973	0930
7	002450995	016682	AA00	197309	0166	821973	0930
8	003529227	009520	AA00	197309	0095	201973	0930
9	003712002	007694	AA00	197309	0076	201973	0930
10	003755953	016682	AA00	197309	0166	821973	0930

- Read data set into Python and parsed columns
- Deleted individuals with “nonsense” entries such as ### or ***
- Deleted individuals with only one working year
- Looped through folder and completed process for all 40 data sets

Pay and Location: Budget and Total Pay

Top Ten Employee Budgets

Location	Budget (Millions of Dollars)
DISTRICT OF COLUMBIA	300
BELLEVILLE	100
KANSAS CITY	75
ALBUQUERQUE	65
ARLINGTON	55
SAINT LOUIS	50
MISSOULA	45
FORT COLLINS	40
PORTLAND	35
BOISE	25

EDA: Pay and Location

The left chart is a box plot titled "Average Pay for Locations with >300 employees (Total Average: \$60,058)". The y-axis is labeled "Average Pay" and ranges from 30k to 100k. The box plot shows a median around \$60,058, with a box from approximately \$58,000 to \$85,000. Whiskers extend from 30k to 100k. Individual data points are plotted as blue dots, with ARLINGTON at the top and MEMPHIS at the bottom.

The right chart is a horizontal bar chart titled "Top Ten Average Pay (locations with over 100 employees)". The x-axis is labeled "Average Pay" and ranges from 0k to 100k. The y-axis lists ten locations. The bars are blue and ordered from highest to lowest average pay.

Location	Average Pay (Approximate)
ARLINGTON	95k
DISTRICT OF COLUMBIA	92k
VALLEJO	90k
HYDROBORO	88k
KANSAS CITY	85k
ATLANTA	82k
GOLDEN	80k
ALBANY	78k
BELTSVILLE	75k
PORTLAND	72k

A histogram titled "Histogram of Percent Change of Pay Over Career". The x-axis is labeled "Percent Change of Pay Over Career" and ranges from 0 to 2000 with major ticks every 250 units. The y-axis is labeled "Frequency" and ranges from 0.0 to 1.2 with major ticks every 0.2 units. The histogram consists of 15 bars of equal width (125 units each). The distribution is right-skewed, with the highest frequency of approximately 1.25 occurring in the first bin (0-125). The frequency decreases as the percent change increases, with the last bin (1875-2000) having a frequency of 0.0.

Percent Change of Pay Over Career (Bin Range)	Frequency
0 - 125	1.25
125 - 250	0.95
250 - 375	0.65
375 - 500	0.45
500 - 625	0.35
625 - 750	0.25
750 - 875	0.18
875 - 1000	0.12
1000 - 1125	0.08
1125 - 1250	0.05
1250 - 1375	0.03
1375 - 1500	0.02
1500 - 1625	0.01
1625 - 1750	0.01
1750 - 1875	0.01
1875 - 2000	0.00

- **Career:** employee that appears in data set more than once
- Career advancement can be measured by base pay and grade variables.
- We defined a variable that tracks an employee's change in pay over their career
 - Used pandas package in python to create pivot tables that organized the basic pay for each individual
 - Calculated percent pay change from the difference in pay between the first and last years worked
 - ~3.2% of data lost due to missing values

Location and Agency

Map based on Longitude (generated) and Latitude (generated). Size shows median of Percent Career Pay Change. Details are shown for Distation and Age. The view is filtered on Latitude (generated) and Longitude (generated). The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only.

- We identified locations and agencies with particularly high median percent change in pay over employee s careers.
- A number of locations with high percent median change in pay were Rural Development Agencies.

- Approximately 12,000 individuals who worked for the Rural Development Agency saw a 9000% increase in pay over their career.
- Indicated the possible existence of an employment program.

Median Percent Change in Pay Over Career by Location

Median Percent Cx
0 13.33%

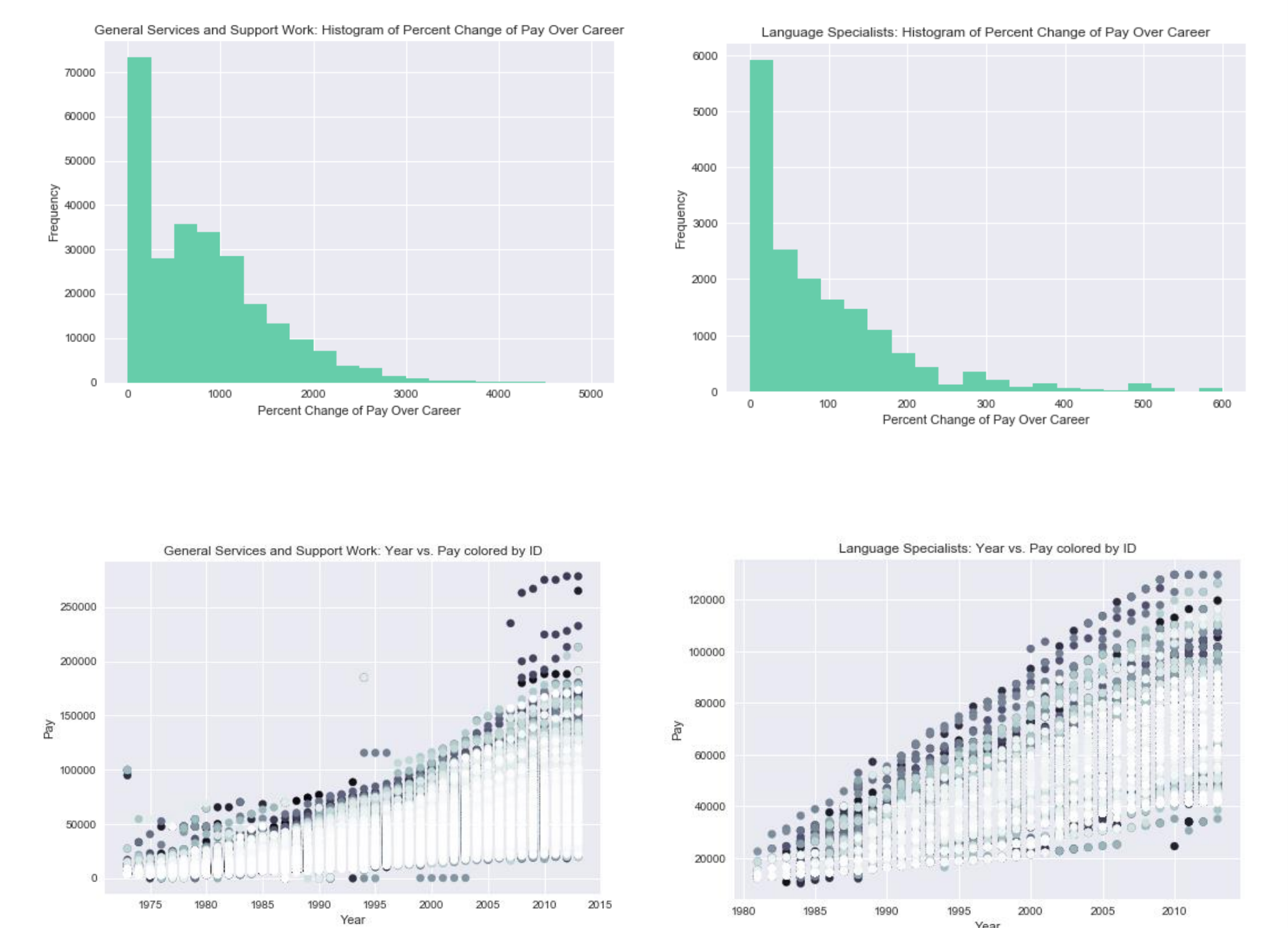
Color shows median of Percent Cancer Pay Change. Size shows median of Percent Cancer Pay Change. The means are labeled by Dotation. The view is filtered on basis of Percent Cancer Pay Change, which includes greater than and/or equal to End Screen 1st value.

Ague. Color shows median of Percent Cancer Pay Change. Size shows median of Percent Cancer Pay Change. The means are labeled by Size. The data is filtered on average of Percent Cancer Pay Change, which means

- Farm Credit Insurance Corporation had a median of approximately 2000 percent increase in pay over career, as compared to the rest of agencies that ranged from 100-1000 percent.
- Identified five locations with significantly higher median percent change in pay over career than the rest of the population.
 - Davis Dam, Big Rapids, Ogalla, Newark Valley, and West Feliciana saw a median percent change in pay over careers of approximately 10,000 percent.

Entry Position

- Employees who started their careers in General Services and Support Work saw a high percent change in pay over their career.
- Compared these graphs with an occupation that saw more normal percent change in pay over careers – Language Specialists.
 - Both histograms displayed a similar trend
 - General Services had a larger range
 - Scatterplot of General Services showed a few individuals with sharp pay increases over their careers
 - Scatter plot of Language Specialists indicated steady pay increase over careers.



Preliminary Conclusions

- Entry position, agency, and geographic location have an effect on career advancement.
- Next Steps
- Investigate the causes behind specific locations, agencies, and entry positions having higher median percent change in pay over careers going forward.
 - Conduct a thorough analysis of all variables to fully understand what is leading to these high values.
 - Examine influence of managers on career advancement.
 - We will need to develop a method for identifying departmental managers and break down the data set into different groups of employees by their manager.

Pandas for Everyone: Python Data Analysis. Pearson Education (US), 2017.

OPM Personnel Documentation Guidance.” *U.S. Office of Personnel Management*, www.opm.gov/policy-data-oversight/data-analysis-documentation/personnel-documentation/.

Cheng, Haiyan. Willamette University Computer Science.

Johnson, Tim. Atkinson Graduate School of Management.