EFFECTS ON EMPLOYMENT RATES & INCOME:

MAJOR CATEGORY, GENDER, GRADUATE DEGREE

STAT 040, Introduction to R and SAS Final Project

Mason Sommers, Olivia Soppa, and Kendall Starcevich

ACKNOWLEDGEMENTS

First, we would like to thank our professor, Dr. Wiranthe Herath of the Drake University Statistics Department, for equipping us with the knowledge to complete this project. It would not have been possible to make these findings without his assistance in:

a. teaching us the proper functions in both R and SAS,
b. strengthening our analytical skills to be able to analyze our findings,

and

c. answering any questions throughout our data collection, proposal, first draft, and final draft processes.

We appreciate all of his valuable guidance and support throughout the semester.

Second, we would like to thank Ben Casselman for making his data public and available for our analysis.

INTRODUCTION

Project Description

We will be using SAS and R programming languages to interpret data based on employment rates and income of different college majors after graduation. We also looked at the effect of gender and obtaining a graduate degree on income and employment rates.

Project Objectives

We wanted to determine if there is a correlation between certain majors and high employment rates. We can also group the majors by category (ex: Law and Public Policy, Health, Humanities and Liberal Arts) and see if there is a correlation between different categories of studies and their employment rates.

We also wanted to determine if there was a difference in the median salaries of men vs. women, and if certain major categories were a majority of one gender.

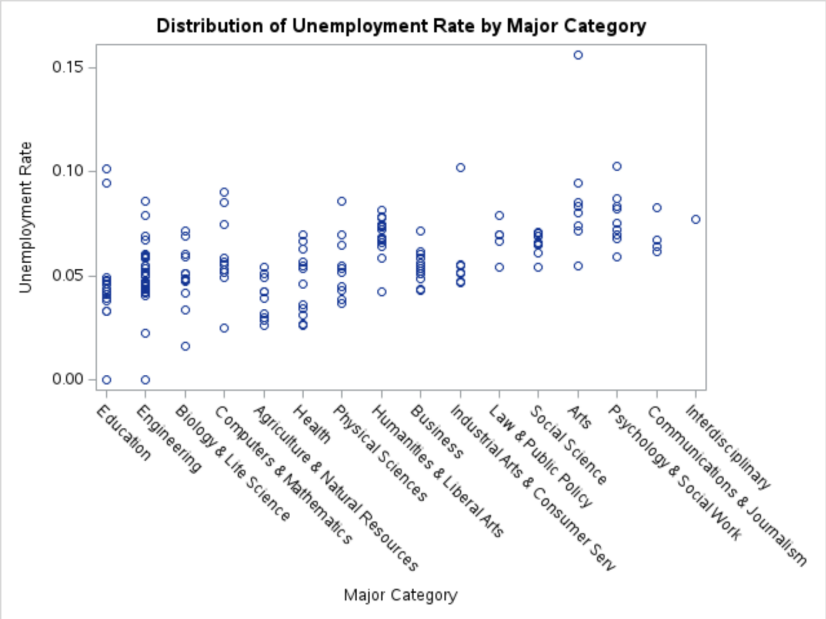
Finally, we wanted to see if obtaining a graduate degree resulted in higher employment rates or salaries, and if this is consistent among all major categories.

Data Description

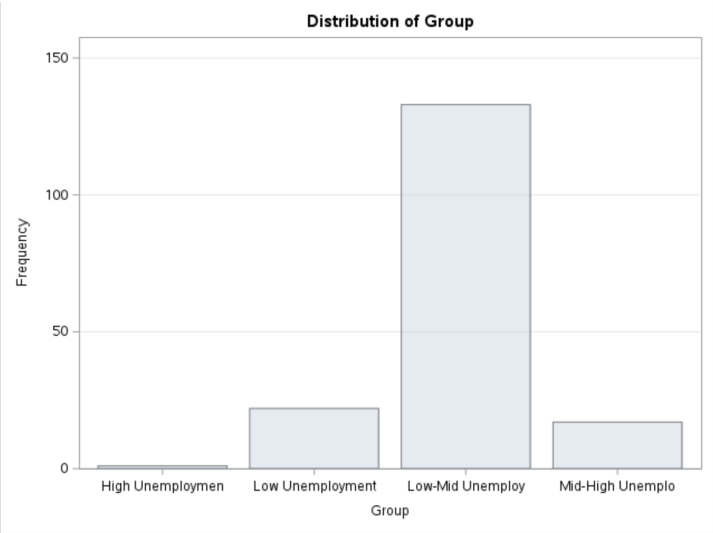
All data is from csv files compiled from the American Community Survey 2010-2012 Public Use Microdata Series. The major categories are from Carnevale et al, "What's It Worth?: The Economic Value of College Majors." Georgetown University Center on Education and the Workforce, 2011.We used three main data files:

1. all-ages.csv
2. recent-grads.csv (ages<28)
3. grad-students.csv (ages 25+)

RESULTS: SAS



This graph is a scatter plot distribution of the unemployment rate based on major categories for all ages. Each data point represents the unemployment rate for a college major categorized based on the y-axis variable, major category. The graph displays the variance of unemployment rate based on major categories (which is more consistent/predictable) and its range as compared to the other categories. The average unemployment rate of the data set is 0.0573554533, so the data points respectively lie above or below that value depending on whether they have a high or low unemployment rate. Education, Agriculture & Natural Resource, and Health major categories most consistently have below average unemployment rates. Arts, Psychology & Social Work and Humanities & Liberal Arts major categories most consistently have above average unemployment rates.



This graph is a representation of the distribution of unemployment rates. The groups are categorized by low unemployment rate (<0.04), low-mid unemployment (0.04-0.08), mid-high unemployment (0.08-0.12) and high unemployment (>0.12). The most frequent occurring unemployment category is low-mid unemployment and the least frequent occurring unemployment category is high unemployment.

Statistical Properties of Unemployment Rate by College Major Data

Obs	Major_category	_TYPE_	_FREQ_	Mean_Unemployment_rate
1	Arts	1	8	0.0876005171
2	Psychology & Social Work	1	9	0.0778670166
3	Interdisciplinary	1	1	0.077268968
4	Humanities & Liberal Arts	1	15	0.0694287074
5	Communications & Journalism	1	4	0.0691245208
6	Law & Public Policy	1	5	0.0678535622
7	Social Science	1	9	0.0656856623
8	Computers & Mathematics	1	11	0.0594369838
9	Industrial Arts & Consumer Serv	1	7	0.0585456723
10		0	173	0.0573554533
11	Physical Sciences	1	10	0.0545406169
12	Business	1	13	0.0544960153
13	Engineering	1	29	0.050630023
14	Biology & Life Science	1	14	0.0499359685
15	Health	1	12	0.047209279
16	Education	1	16	0.0467619641
17	Agriculture & Natural Resources	1	10	0.0395691804

This table is a numerical representation of the mean unemployment rate by major category in descending order. The highest average unemployment rate major category is Arts, and the lowest is Agriculture & Natural Resources. Observation 10 is the average unemployment rate of all of the data, therefore observations 1-9 are above average unemployment and observations 11-17 are below average.

Highest Unemployment Rates of Employment by College Major Data

Obs	Major	Major_category	Employed	Unemployed	Unemployment_rate
1	MISCELLANEOUS FINE ARTS	Arts	6431	1190	0.156147487
2	CLINICAL PSYCHOLOGY	Psychology & Social Work	5128	587	0.102712161
3	MILITARY TECHNOLOGIES	Industrial Arts & Consumer Serv	1650	187	0.101796407
4	SCHOOL STUDENT COUNSELING	Education	1492	169	0.101745936

Lowest Unemployment Rates of Employment by College Major Data

Obs	Major	Major_category	Employed	Unemployed	Unemployment_rate
1	PHARMACOLOGY	Biology & Life Science	3481	57	0.016110797
2	EDUCATIONAL ADMINISTRATION AND SUPERV	Education	3113	0	0
3	GEOLOGICAL AND GEOPHYSICAL ENGINEERIN	Engineering	4120	0	0

These tables display the highest (top) and lowest (bottom) unemployment rates by college major. Misc. Fine Arts, Clinical Psychology, Military Technologies, and School Student Counseling are the highest unemployed majors with respective percentages of 15.6%, 10.3%, 10.2% and 10.2%. Misc. Fine Arts was an outlier of the dataset having a 0.05% higher unemployment rate than the rest of the data. Pharmacology, Educational Admin and Supervision and Geological and Geophysical Engineering are the lowest unemployed majors with respective percentages of 0.02%, 0% and 0% (the last two having 0 occurrences of unemployment).

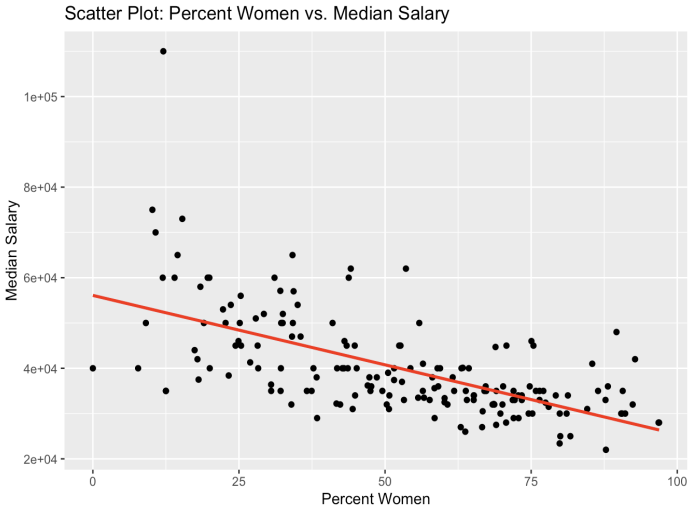
Statistical Properties of Unemployment Rate by College Major Data

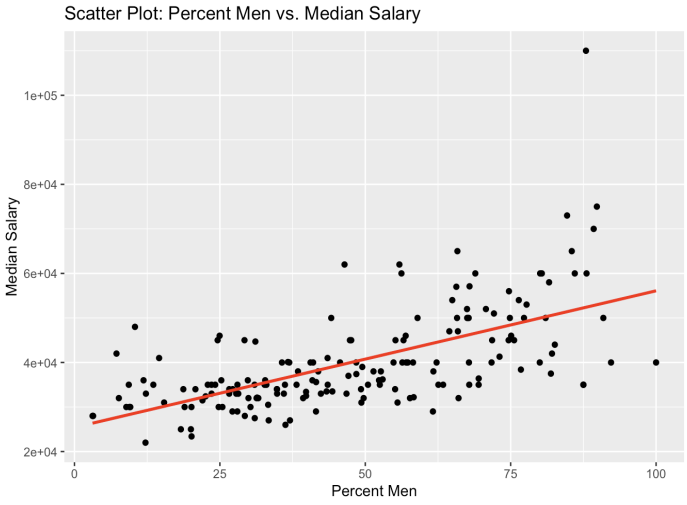
The MEANS Procedure

Analysis Variable : Unemployment_rate						
Major_category	N Obs	N	Mean	Std Dev	Minimum	Maximum
Agriculture & Natural Resources	10	10	0.0395692	0.0100233	0.0261471	0.0543413
Arts	8	8	0.0876005	0.0300967	0.0547192	0.1561475
Biology & Life Science	14	14	0.0499360	0.0138958	0.0161108	0.0715975
Business	13	13	0.0544960	0.0076058	0.0432683	0.0713537
Communications & Journalism	4	4	0.0691245	0.0095035	0.0619168	0.0830048
Computers & Mathematics	11	11	0.0594370	0.0181721	0.0249004	0.0902642
Education	16	16	0.0467620	0.0232375	0	0.1017459
Engineering	29	29	0.0506300	0.0157612	0	0.0859911
Health	12	12	0.0472093	0.0157655	0.0262916	0.0700098
Humanities & Liberal Arts	15	15	0.0694287	0.0095429	0.0425051	0.0813481
Industrial Arts & Consumer Serv	7	7	0.0585457	0.0193734	0.0469030	0.1017964
Interdisciplinary	1	1	0.0772690	.	0.0772690	0.0772690
Law & Public Policy	5	5	0.0678536	0.0090703	0.0540356	0.0792169
Physical Sciences	10	10	0.0545406	0.0153799	0.0367263	0.0860215
Psychology & Social Work	9	9	0.0778670	0.0127381	0.0593759	0.1027122
Social Science	9	9	0.0656857	0.0052784	0.0543988	0.0710569

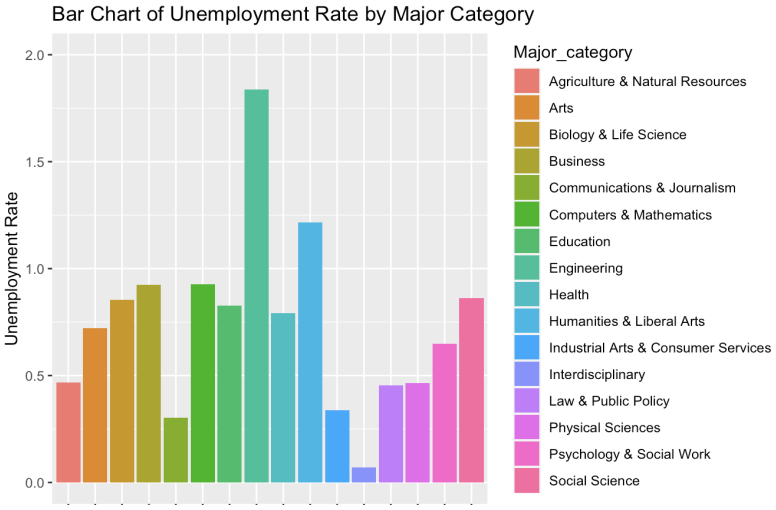
This table is a representation of the statistical properties of the unemployment rate organized by college major category. It includes the number of observations (N Obs), mean, standard deviation (Std Dev), minimum and maximum. The minimum and maximum can be examined to demonstrate the range of the unemployment rate (how much it varies within each field).

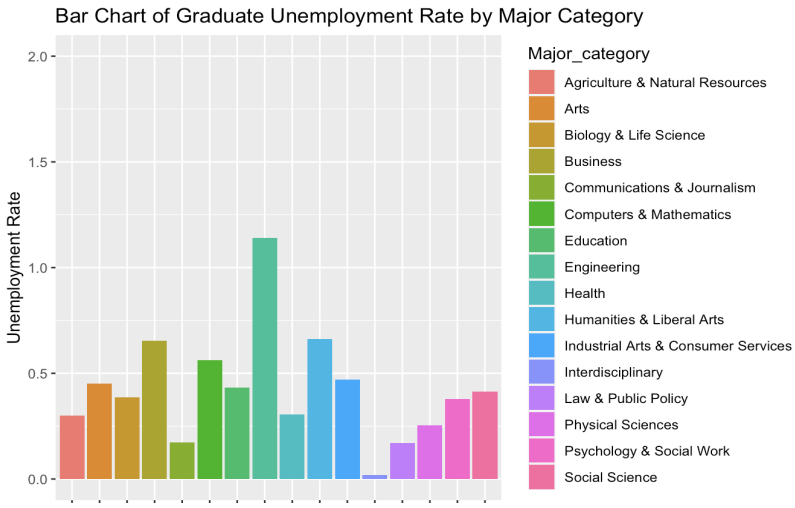
RESULTS: R



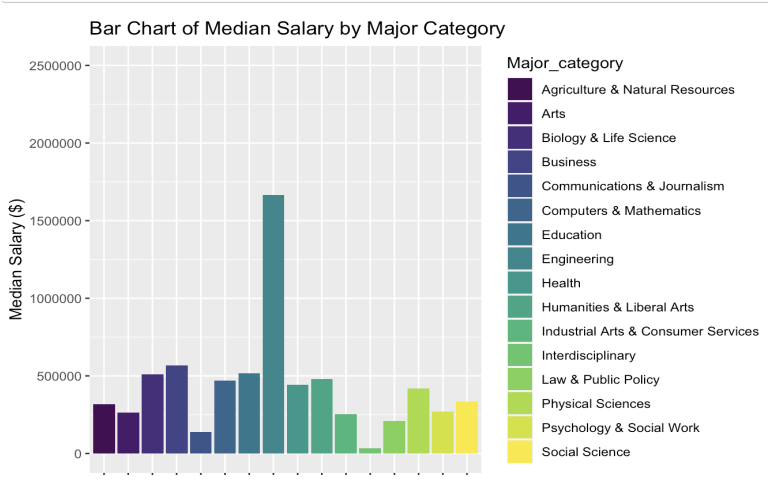


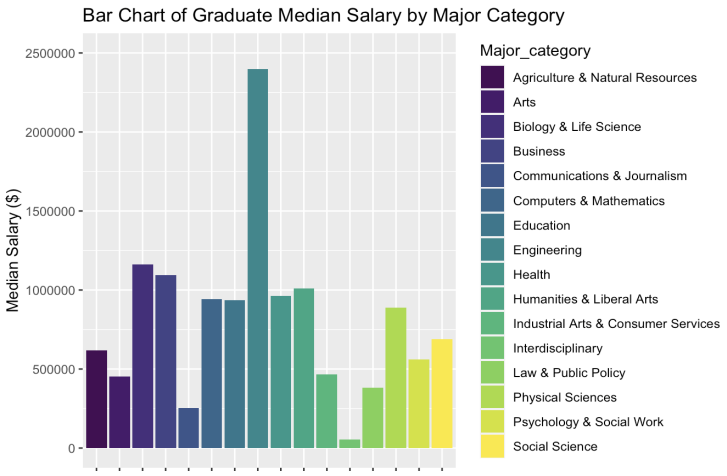
These two graphs are scatter plots with regression lines that represent the correlation of the percentage of men and women in a major compared to the average median salary in that major. Each point is a different major with the x-axis representing the percentage of men or women and the y-axis represents the median salary for that major. These two graphs represent the opposite of the other. In the graph representing women you can see that the graph has a medium negative correlation of -0.6186898. This means as the percentage of women in a major rises the median salary decreases. Since the graph for men is the opposite it has a positive median correlation of 0.6186898. This means as the percentage of men in a major rises the medium salary increases.





These two graphs are bar charts that represent the unemployment rate for each major category. The first graph represents data for only non-graduates and the second graph represents data for only graduates. On the x-axis of both graphs each color represents a different major category and on the y-axis is the unemployment rate for that major category. It's clear that in both of the graphs the major category with the highest unemployment rate is engineering and the major category with the lowest employment rate is interdisciplinary. You can also tell by looking at both graphs that graduate students have a much lower unemployment rate for every major compared to the unemployment rate of non-graduate students.





These two graphs are bar graphs that represent the median salary for every major category. The first graph represents data for only non-graduates and the second graph represents data for only graduates. On the x-axis of both graphs each color represents a different major category and on the y-axis is the median salary for that major category. By looking at the graph you can see engineering has the highest average median salary and interdisciplinary has the lowest median salary for both graduates and non-graduates. It's also very clear that there is a big difference in the amount of money made with graduates in all fields making significantly more.

CONCLUSION

Major Categories

Education, Agriculture & Natural Resource, and Health major categories most consistently have below average unemployment rates. Arts, Psychology & Social Work and Humanities & Liberal Arts major categories most consistently have above average unemployment rates.

Gender

There is a correlation between gender and median salary. The higher the number of men in the major category, the higher the median salary.

Graduate Degrees

For every major category, the employment rate and median salaries are higher for people with a graduate degree.

REFERENCES

The csv files are from Ben Casselman's Github: *BenCasselman*, and we found them on Github's FiveThirtyEight. Furthermore, all data from the csv files are compiled from the American Community Survey 2010-2012 Public Use Microdata Series. The major categories are from Carnevale et al, "What's It Worth?: The Economic Value of College Majors." Georgetown University Center on Education and the Workforce, 2011.

Our sources for the R and SAS code are from Professor Perath's STAT 040 lectures and demos.