Will You Find a Job After College?

Addis Pimentel

2/25/24

Contents

Getting Started

As a college student, I think we all wonder about our life post graduation. “Will I get the job I always wanted?”, “Will I be able to pay off my debt?”, “Will I make a lot of money?”, “Will I even get a job at all?” Most college students think about these questions almost everyday, especially for students who are paying for college or taking out loans, it’s important to have an idea of what to expect after college. For curious college students and potential college students, we will be looking at unemployment rate and median salary for different majors.

About the Data Set

The data-set we will be using in this analysis is the “college_all_ages” data-set from the “fivethirtyeight” package. This data-set represents majors for college graduates of all ages along with 11 other variables:

- major_code : Major code
- major : Major description
- major_category : Category of major
- total : Total number of people with major
- employed : Number employed
- employed_fulltime_yearround : Employed at least 50 weeks and at least 35 hours
- unemployed : Number unemployed
- unemployment_rate : Unemployed/(Unemployed + Employed)
- p25th : 25th percentile of earnings
- median : Median earnings of full-time, year-round workers
- p75th : 75th percentile of earnings

Source for The ‘college_all_ages’ Data Set

Analyzing the Unemployment Rate Among College Graduates by Category of Major

Here we are visualizing the relationship between unemployment rate and the different categories of majors within this data-set. Since we are comparing a numeric variable with a categorical variable of more than 6 categories, a Ridgeline Plot is the best choice for visualizing this relationship. On our x-axis we have the unemployment rate in percentage for each category of major. By analyzing our plot, we can see

The Top 5 Categories of Majors with the Highest Unemployment Rates are:

- Arts
- Industrial Arts & Consumer Services
- Psychology & Social Work
- Education
- Computers & Mathematics

The Top 3 Categories of Majors with the Largest Variability in Unemployment Rates are:

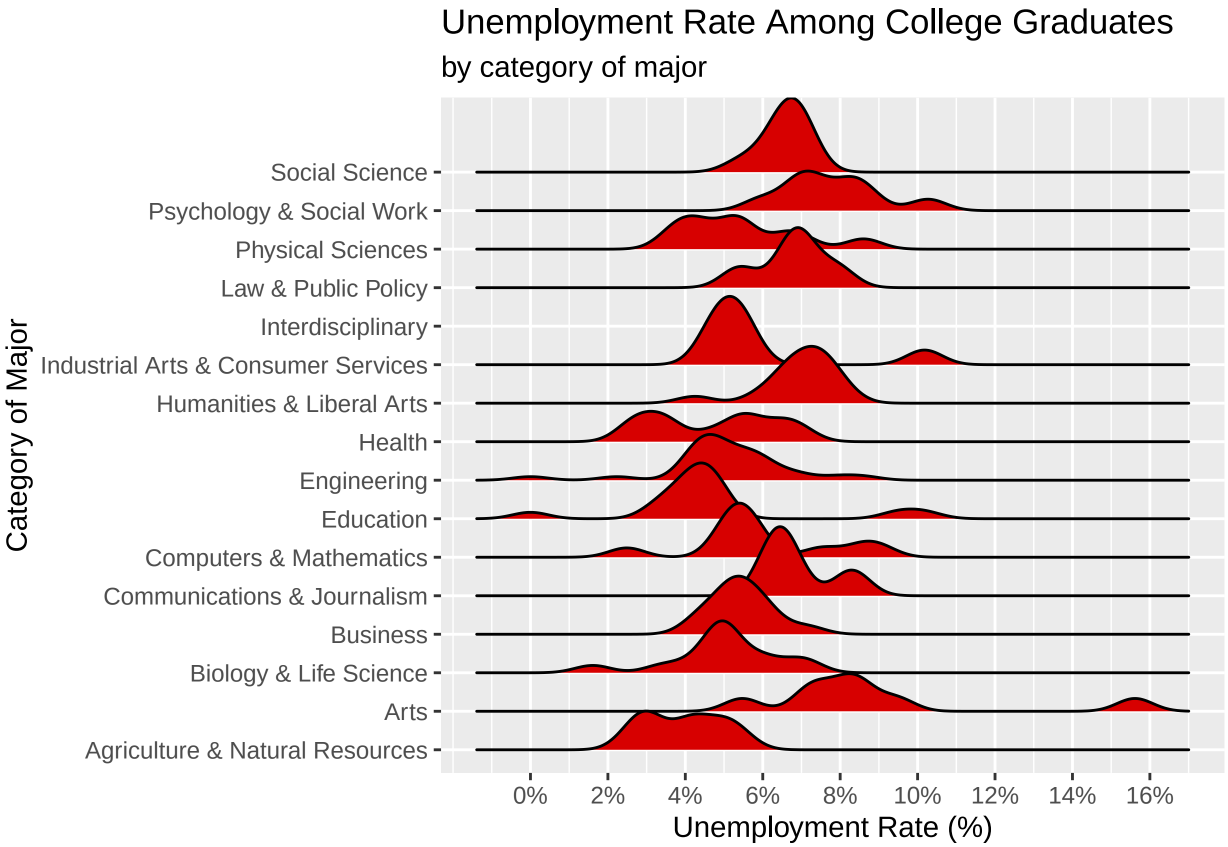
- Arts
- Education
- Engineering

What does this mean?

For college graduates with majors in 'The Top 5 Categories (of Majors) with the Highest Unemployment Rates', the overall pattern from our data suggests that unemployment is higher among college graduates of these 5 categories (Arts, Industrial Arts & Consumer Services, Psychology & Social Work, Education, and Computers & Mathematics). And for college graduates in the 'Top 3 Categories of Majors with the Largest Variability in Unemployment Rates' the overall pattern from our data suggests there is a larger variability in unemployment rates among graduates of these 3 categories (Arts, Education, Engineering) compared to graduates of other major categories. The result of this data could be affected by other unknown variables, therefore a proper inference cannot be made.

Ridgeline Plot

```r
ggplot(data = college_all_ages, aes(x = unemployment_rate, y = as.factor(major_category))) +
  geom_density_ridges(color = "black", fill = "#d70000", linewidth = .5)+
  labs(x = "Unemployment Rate (%)", y = "Category of Major",
       title = "Unemployment Rate Among College Graduates",
       subtitle = "by category of major")+
  scale_x_continuous(breaks = seq(0,18, by = 2),labels = c("0%","2%","4%","6%","8%", "10%",
                      "12%", "14%","16%","18%"))
```



More about the code

In the above Ridgeline Plot code, we are mapping the unemployment rate variable to our x-axis, with major categories as our categorical variable on our y-axis, this is within our aesthetic function that is within our ggplot function. Our geom_density_ridges function contains the color for the lines of our Ridgeline Plot, in this case I chose black as the outline with a line-width of .5 for our Ridgeline plot and "d70000" as our fill color, this is a color code for a shade of red. Our labels for our plot give some clarity on the variables of interest and the relationship we are interested in visualizing, in this case the we are interested in relationship between unemployment rate and major categories.

Analyzing the Median Salary of Full-Time Working College Graduates by Category of Major.

Here we are Visualizing the relationship between the median salary of a full-time working college graduate and the different categories of majors. We are comparing a numeric variable (median salary) and a categorical variable (major_category), similar to our Ridgeline Plot for unemployment rate and major. On our x-axis we have our median salary which is increasing at an interval of $10,000, starting at $30,000 and ranging up until $130,000. By analyzing our plot, we can see

The Top 5 Categories of Majors with the Highest Median Salary are:

- Engineering
- Health
- Computers & Mathematics
- Physical Sciences
- Business

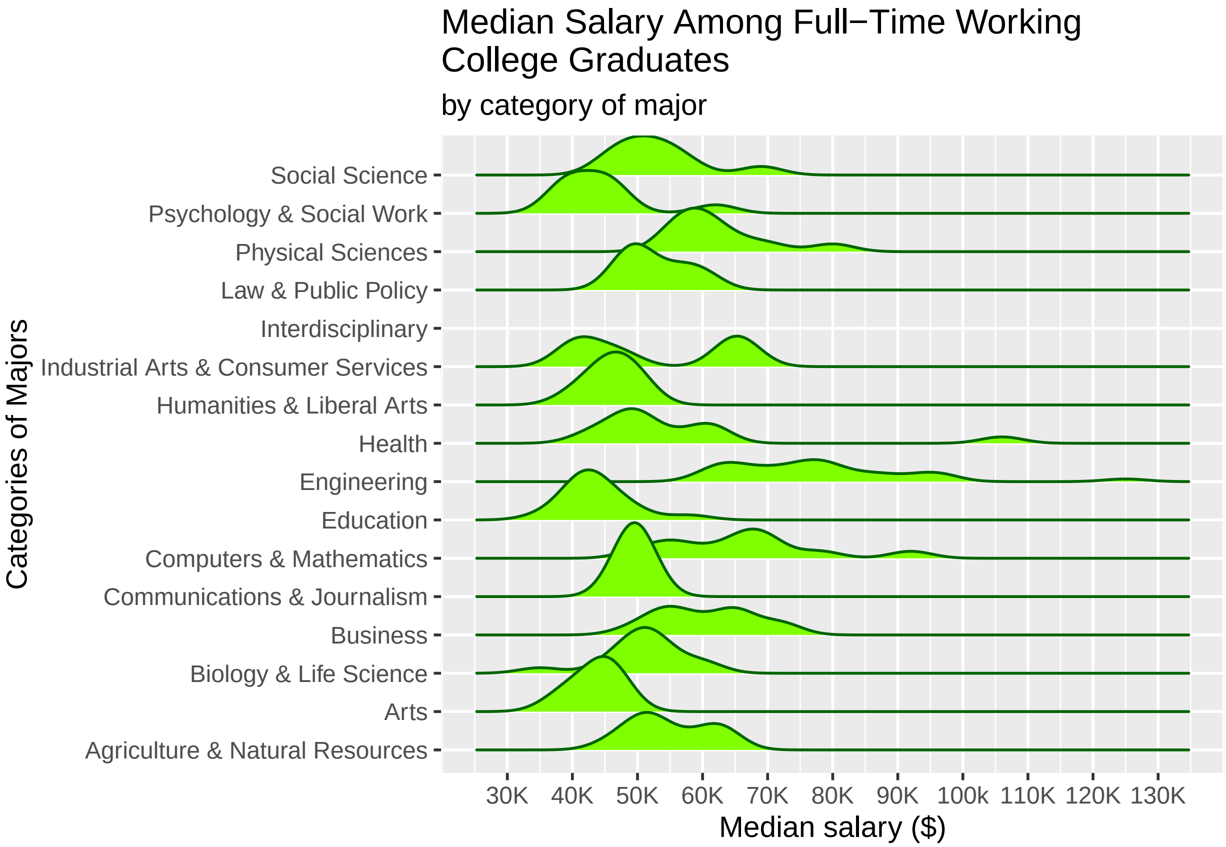
The Top 3 Categories of Majors with the Largest Variability in Median Salary are:

- Engineering
- Health
- Computers & Mathematics

What does this mean? For college graduates with majors in 'The Top 5 Categories (of Majors) with the Highest Median Salary', the overall pattern from our data suggests that the median salary is higher among college graduates of these 5 categories (Engineering, Health, Computers & Mathematics, Physical Sciences, and Business). And for college graduates in the 'with the Largest Variability in Median Salary' the overall pattern from our data suggests there is a larger variability in unemployment rates among graduates of these 3 categories (Engineering, Health, and Computers & Mathematics) compared to graduates of other major categories. The result of this data could be affected by other unknown variables, therefore a proper inference cannot be made.

Ridgeline Plots

```r
ggplot(data = college_all_ages, aes(x = median, y = as.factor(major_category))) +
  geom_density_ridges(color = "darkgreen", fill = "chartreuse", linewidth = .5)+
  labs(x = "Median salary ($)", y = "Categories of Majors",
       title = "Median Salary Among Full-Time Working\nCollege Graduates",
       subtitle = "by category of major")+
  scale_x_continuous(breaks = seq(30000,130000, by =10000),
                     labels = c("30K","40K","50K","60K", "70K",
                      "80K", "90K","100k","110K","120K","130K"))
```



More about the code

In the above Ridgeline Plot code, we are mapping the median variable to our x-axis, with major categories as our categorical variable on our y-axis, this is within our aesthetic function that is within our ggplot function. Our geom_density_ridges function contains the color for the lines of our Ridgeline Plot, in this case I chose dark-green as the outline with a line-width of .5 for our Ridgeline plot and "chartreuse" as our fill color. Our labels for our plot give some clarity on the variables of interest and the relationship we are interested in visualizing, in this case the we are interested in relationship between median salary and major categories. Our x-axis labels are increasing by $10,000 and written as "10K, 20K, 30K. . . " for visual clarity.