Pricilla Nakyazze 607-Week 1 Assignement

2025-08-26

Introduction

Choosing a college major is more than just following one's passion. It can have a measurable impact on future earnings. Even among closely related fields, significant differences in income potential exist. For instance, actuarial science majors tend to out-earn accounting majors, and public policy majors see better earnings outcomes than history majors. Interestingly, vocational fields like court reporting may offer better returns than more traditional majors like criminology. While earning a college degree does not guarantee economic success, data clearly shows that choosing the right major can improve the odds of financial stability.

The full Article on choince of major's impact on employment can be found here: https://fivethirtyeight.com /features/the-economic-guide-to-picking-a-college-major/

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```
install.packages("readr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(dplyr)
library(readr)
library(ggplot2)
```

##1.1 Load the college majors-all ages data into a data frame and preview.

```
majors_data <- read_csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/college-majors/a
```

```
## Rows: 173 Columns: 11
## -- Column specification --------------------------------------------------------
```

```
## Delimiter: ","
## chr (2): Major, Major_category
## dbl (9): Major_code, Total, Employed, Employed_full_time_year_round, Unemplo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(majors_data)
```

```
## Rows: 173
## Columns: 11
## $ Major_code                    <dbl> 1100, 1101, 1102, 1103, 1104, 1105, 1106~
## $ Major                         <chr> "GENERAL AGRICULTURE", "AGRICULTURE PROD~
## $ Major_category                <chr> "Agriculture & Natural Resources", "Agri~
## $ Total                         <dbl> 128148, 95326, 33955, 103549, 24280, 794~
## $ Employed                      <dbl> 90245, 76865, 26321, 81177, 17281, 63043~
## $ Employed_full_time_year_round <dbl> 74078, 64240, 22810, 64937, 12722, 51077~
## $ Unemployed                    <dbl> 2423, 2266, 821, 3619, 894, 2070, 264, 2~
## $ Unemployment_rate             <dbl> 0.02614711, 0.02863606, 0.03024832, 0.04~
## $ Median                        <dbl> 50000, 54000, 63000, 46000, 62000, 50000~
## $ P25th                         <dbl> 34000, 36000, 40000, 30000, 38500, 35000~
## $ P75th                         <dbl> 80000, 80000, 98000, 72000, 90000, 75000~
```

Select the Major column and return it as a vector

```
majors <- majors_data |>
  select(Major) |>
  pull()
head(majors, n=12)
```

```
##  [1] "GENERAL AGRICULTURE"
##  [2] "AGRICULTURE PRODUCTION AND MANAGEMENT"
##  [3] "AGRICULTURAL ECONOMICS"
##  [4] "ANIMAL SCIENCES"
##  [5] "FOOD SCIENCE"
##  [6] "PLANT SCIENCE AND AGRONOMY"
##  [7] "SOIL SCIENCE"
##  [8] "MISCELLANEOUS AGRICULTURE"
##  [9] "ENVIRONMENTAL SCIENCE"
## [10] "FORESTRY"
## [11] "NATURAL RESOURCES MANAGEMENT"
## [12] "ARCHITECTURE"
```

Identifies the majors that contain "AGRICULTURE"

```
str_view(majors, "AGRICULTURE")
```

```
## [1] | GENERAL <AGRICULTURE>
## [2] | <AGRICULTURE> PRODUCTION AND MANAGEMENT
## [8] | MISCELLANEOUS <AGRICULTURE>
```

students must approach their college decisions with care. Choosing a major with stronger labor market outcomes not only boosts earning potential but also reduces the risk of graduating into low-income brackets. The worst-case scenario? Ending up in the bottom 25% of earners, where attending college may not have paid off financially. Psychology is considered a major with lor return on Investment. Business majors are

often influenced by many variables so that even though unemployment is high the number of students that are gainfully employed after graduating is high as well.

Get top 4 majors with most unemployed graduates.

```
top4_unemployed_majors <- majors_data |>
  dplyr::arrange(desc(Unemployed)) |>
  head(4) |>
  dplyr::arrange(Unemployed)

# View the result
print(top4_unemployed_majors)
```

```
## # A tibble: 4 x 11
##   Major_code Major         Major_category  Total Employed Employed_full_time_y~1
##        <dbl> <chr>         <chr>           <dbl>    <dbl>                   <dbl>
## 1       6201 ACCOUNTING    Business       1.78e6  1335825                 1095027
## 2       5200 PSYCHOLOGY    Psychology & ~ 1.48e6  1055854                  736817
## 3       6200 GENERAL BUSI~ Business       2.15e6  1580978                 1304646
## 4       6203 BUSINESS MAN~ Business       3.12e6  2354398                 1939384
## # i abbreviated name: 1: Employed_full_time_year_round
## # i 5 more variables: Unemployed <dbl>, Unemployment_rate <dbl>, Median <dbl>,
## #   P25th <dbl>, P75th <dbl>
```

library(dplyr)

Get top 10 majors with most unemployed, showing specific columns, sorted in ascending order

```
top10_unemployed_majors <- majors_data |>
  arrange(desc(Unemployed)) |>
  slice_head(n = 10) |>
  arrange(Unemployed) |>
  select(Major, Total, Employed, Employed_full_time_year_round, Unemployed)

# View the result
print(top10_unemployed_majors)
```

```
## # A tibble: 10 x 5
##    Major                          Total Employed Employed_full_time_y~1 Unemployed
##    <chr>                          <dbl>    <dbl>                  <dbl>      <dbl>
##  1 BIOLOGY                       8.39e5   583079                 422788      36757
##  2 GENERAL EDUCATION             1.44e6   843693                 591863      38742
##  3 POLITICAL SCIENCE AND GOVE~ 7.49e5   541630                 421761      40376
##  4 MARKETING AND MARKETING RE~ 1.11e6   890125                 704912      51839
##  5 ENGLISH LANGUAGE AND LITER~ 1.10e6   708882                 482229      52248
##  6 COMMUNICATIONS                9.88e5   790696                 595739      54390
##  7 ACCOUNTING                    1.78e6  1335825                1095027      75379
##  8 PSYCHOLOGY                    1.48e6  1055854                 736817      79066
##  9 GENERAL BUSINESS              2.15e6  1580978                1304646      85626
## 10 BUSINESS MANAGEMENT AND AD~ 3.12e6  2354398                1939384     147261
## # i abbreviated name: 1: Employed_full_time_year_round
```

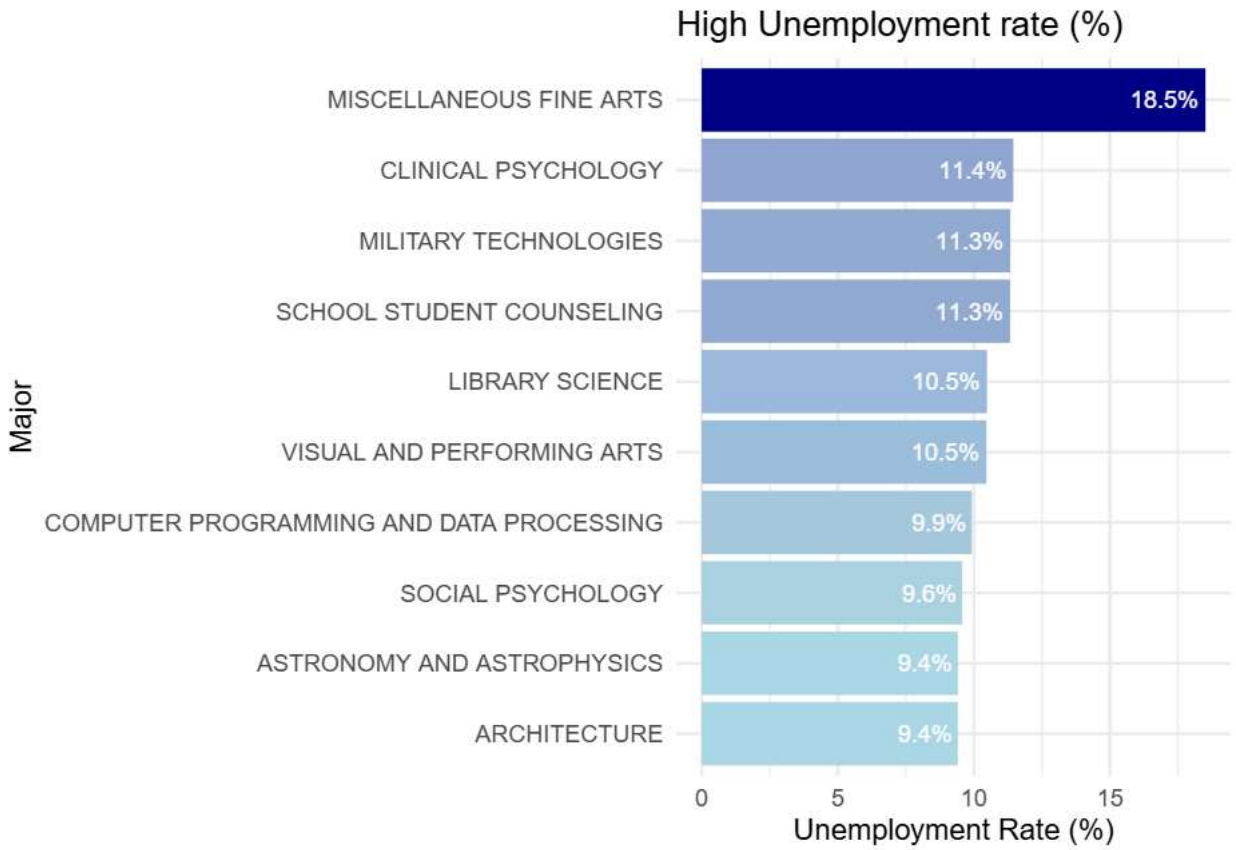
Get top 4 majors with highest employed graduates, sorted in ascending order

```
top4_employed_majors <- majors_data |>
  dplyr::arrange(desc(Employed)) |>
  head(4) |>
  dplyr::arrange(Employed)

# View the result
print(top4_employed_majors)
```

```
## # A tibble: 4 x 11
##   Major_code Major         Major_category  Total Employed Employed_full_time_y~1
##        <dbl> <chr>         <chr>           <dbl>    <dbl>                   <dbl>
## 1       6107 NURSING       Health         1.77e6  1325711                  947546
## 2       6201 ACCOUNTING    Business       1.78e6  1335825                 1095027
## 3       6200 GENERAL BUSI~ Business       2.15e6  1580978                 1304646
## 4       6203 BUSINESS MAN~ Business       3.12e6  2354398                 1939384
## # i abbreviated name: 1: Employed_full_time_year_round
## # i 5 more variables: Unemployed <dbl>, Unemployment_rate <dbl>, Median <dbl>,
## #   P25th <dbl>, P75th <dbl>
```

A more accurate way to tell the return on investment on a course is by percentage of students employed.

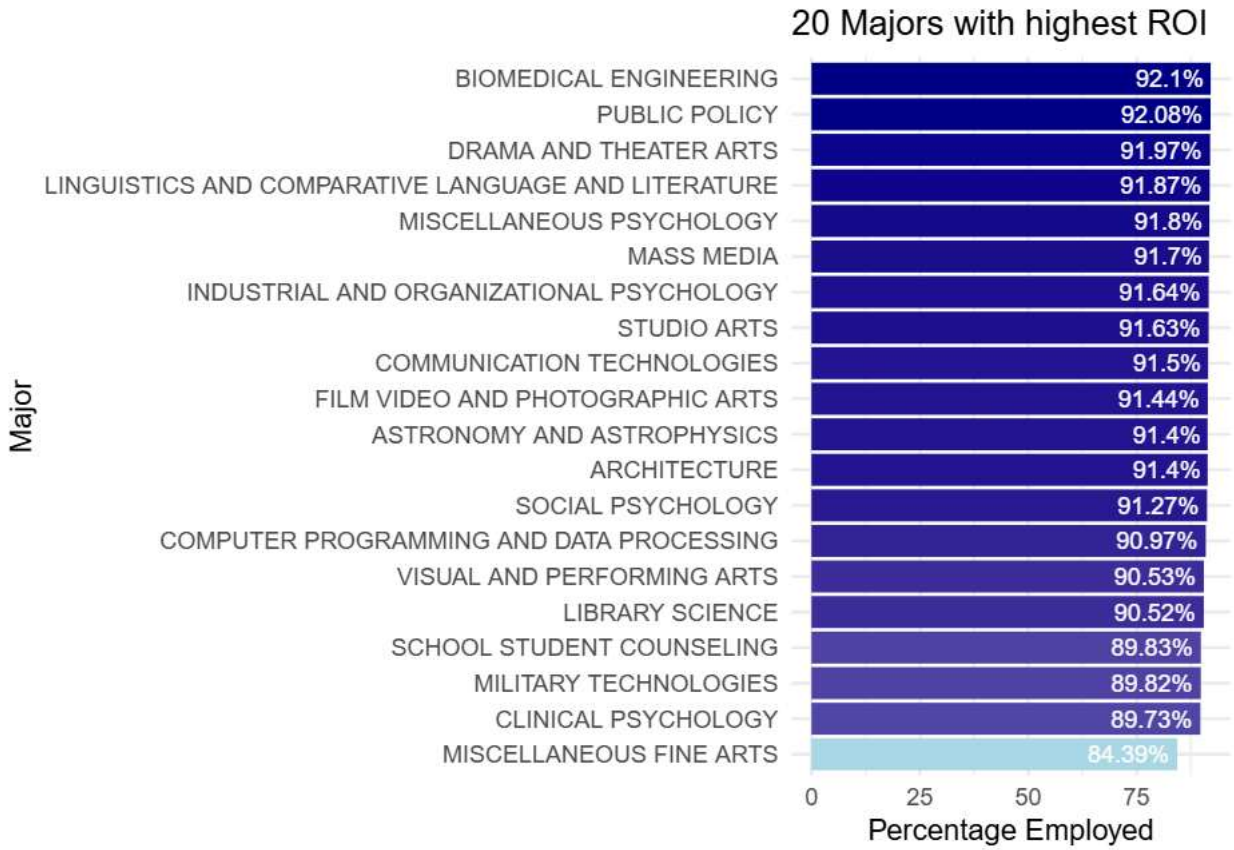
```
# Calculate unemployment percentage and get top 10 majors with highest % unemployed
top10_unemployed_pct <- majors_data |>
  mutate(Unemployment_Percent = (Unemployed / Employed) * 100) |>
  arrange(desc(Unemployment_Percent)) |>
  slice_head(n = 10) |>
  arrange(Unemployment_Percent)  # Ascending order for plotting


# Plot
ggplot(top10_unemployed_pct, aes(x = reorder(Major, Unemployment_Percent),
                                 y = Unemployment_Percent,
                                 fill = Unemployment_Percent)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = sprintf("%.1f%%", Unemployment_Percent)),
            hjust = 1.1, color = "white", size = 3) +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  coord_flip() +
  labs(
    title = "High Unemployment rate (%)",
    x = "Major",
    y = "Unemployment Rate (%)"
  ) +
  theme_minimal() +
  theme(legend.position = "none")  # optional: remove legend if not needed
```




```
##PercentageEmployed = Employed / (Employed + Unemployed) * 100



# Create summary table
employment_summary <- majors_data |>
  mutate(PercentageEmployed = round(Employed / (Employed + Unemployed) * 100, 2)) |>
  select(Major, Employed, Unemployed, PercentageEmployed)

# View the result
print(employment_summary)
```

```
## # A tibble: 173 x 4
##    Major                                Employed Unemployed PercentageEmployed
##    <chr>                                   <dbl>      <dbl>               <dbl>
##  1 GENERAL AGRICULTURE                     90245       2423                97.4
##  2 AGRICULTURE PRODUCTION AND MANAGEMENT   76865       2266                97.1
##  3 AGRICULTURAL ECONOMICS                  26321        821                97.0
##  4 ANIMAL SCIENCES                         81177       3619                95.7
##  5 FOOD SCIENCE                            17281        894                95.1
##  6 PLANT SCIENCE AND AGRONOMY              63043       2070                96.8
##  7 SOIL SCIENCE                             4926        264                94.9
##  8 MISCELLANEOUS AGRICULTURE                6392        261                96.1
##  9 ENVIRONMENTAL SCIENCE                   87602       4736                94.9
## 10 FORESTRY                                48228       2144                95.7
## # i 163 more rows
```

```r
# Filter the 20 lowest percentage employed majors
lowest_20 <- employment_summary |>
  arrange(PercentageEmployed) |>
  slice_head(n = 20)
# Plot
ggplot(lowest_20, aes(x = reorder(Major, PercentageEmployed),
                      y = PercentageEmployed,
                      fill = PercentageEmployed)) +
  geom_col() +
  coord_flip() +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  geom_text(aes(label = paste0(PercentageEmployed, "%")),
            hjust = 1.1, color = "white", size = 3) +
  labs(
    title = "20 Majors with highest ROI",
    x = "Major",
    y = "Percentage Employed"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
```



Conclusions

Despite growing doubts about the value of a college degree, research shows that a bachelor's degree remains a worthwhile investment overall. In fact, a recent study by the Federal Reserve Bank of New York finds that the financial return on a college degree is near its historical peak,even after accounting for rising tuition costs.

That said, students must approach their college decisions with care. Choosing a major with stronger labor

market outcomes not only boosts earning potential but also reduces the risk of graduating into low-income brackets. The worst-case scenario? Ending up in the bottom 25% of earners, where attending college may not have paid off financially.

Recommendations.

Evaluate return on investment when selecting a major. Use non conventional means to get an education like mentoring, internships and apprenticeships.

Research earnings data by major, ideally from reliable sources like the U.S. Census or labor market studies and Pay scale

Schools should be transparent about the return on investment for different majors and policies supported that make this kind of data more accessible to prospective students.

By combining informed decision-making with personal interests, students can pursue degrees that offer both fulfillment and financial stability.

RESOURCES

CHATGPT

https://fivethirtyeight.com/features/the-economic-guide-to-picking-a-college-major/

https://www.geeksforgeeks.org/r-language/graph-plotting-in-r-programming/