Work Life Balance vs. Earnings: Does More Work Mean More Pay & How Education Affects Future Financial Situation

Nur Ameera Sabrina Mahadir & Siyona Behera

2025-05-07

Research Topic: Work Life Balance vs. Earnings

This research focuses on the relationship between work hours and yearly earnings. Our group's goal is to investigate whether working more hours actually leads to higher pay, and how this relationship changes across education levels, major job groups (industries), and demographics such as gender and educational background. This question is relevant for students in college or high school to plan their future careers, or for workers to evaluate job demands, and for employers to think about the fair compensation for their employees.

Research Questions

Our main objective in this project is to understand fully whether it is true that "More Work Means More Pay?" Therefore, we aim to explore whether longer work hours result in higher income. Other than that, we also consider how factors like education level and job type affect this relationship. We hypothesize that while moderate increases in working hours may improve earnings, excessive hours might also diminish returns as the employee might have to face burnout, producing low-quality work and other mental health issues that would negatively affect the company they are working for.

Provenance of Our Data

Data Sources-Kaggle:

We used 2 data sets that we found on Kaggle. Kaggle is a subsidiary of Google, which ensures it follows strong security, data integrity, and privacy practices. It is a well-respected platform in the data science community, while users should still be able to critically evaluate other people's datasets' sources, which makes it a trustworthy and educational resource.

1. Adult Income Data Set: Adult Census Income

This data set, which is shared on Kaggle by Ishaan Gupta, is adapted from the UCI Machine Learning Repository, includes over 32,000 entries from the U.S Census. It contains demographic information and is often used to predict whether someone earns more than $450K a year. It includes information like age, education level, hours worked per week, occupation, and income bracket. This data set helped our group to explore how factors like work hours and education relate to one's income. Since this data set is based on actual census data and is widely used in many data science projects, we felt confident and safe to use it for our analysis.

2. College Majors Data Set: Economic Value of College Majors

This data set is from Kaggle and was uploaded by Sahil Shringi. It originally comes from the American Community Survey and was featured by FiveThirtyEight, which made this data set more credible. This data set includes details about different college majors and one's average median income, major group, and gender background. This information gave us insight into how one's college major could affect their future earnings. We agreed to choose this data set as our supplementary data set because it would strengthen our research findings, and it also connects well with our thesis about careers and incomes.

FAIR Principle

Our project report follows the FAIR principle, especially Findable and Reusable. We utilized data sets from Kaggle that are readily available and publicly accessible, ensuring our cleaned data sets and all of our code (for visualizations and analysis) are stored in a shared GitHub repository. This could enable everyone, including the professors, viewers, or anyone, to view, download, or analyze our work without the need to ask for permission from the owner of the project. Other than that, we also made sure that the code is straightforward, with comments and clear variable names, so it will ease the viewers to follow or build on the code for their preferences. We intend to do so because we want our work could actually be reused, and not just submitted once and forgotten. By producing a project that is reusable, accessible, findable, and interoperable, we have learned so much from other people's shared work, and we wanted to give back in the same way to others.

Data Exploration

Before creating any visualizations, we started by making a summary table to the relationship between how many hours people work on average and how much they typically earn, based on their major group. At first glance, it might seem like the more you work, the more you get paid — but the data shows that it's not that straightforward. For example, people in Law & Public Policy work the longest hours (about 48 hours per week) but actually have lower median earnings compared to fields like Engineering and Business, where people work fewer hours but make more money. This suggests that working more hours doesn't always guarantee higher pay. It depends a lot on the field you're in. This table helped us realize that both education and job type play a big role in whether working more actually leads to earning more. The table 1 is as below:

Table 1: Summary of Weekly Work Hours by Income Group

income	count	min	Q1	median	Q3	max	mad	mean	sd
<=50K	24720	1	35	40	40	99	2.97	38.84	12.32
>50K	7841	1	40	40	50	99	7.41	45.47	11.01

Other than that, we also made another summary table that emphasized explore basic statistics on the number of hours people work each week, based on their education level. We included values like the count, minimum, maximum, median, quartiles, mean, and standard deviation to help us see how work hours are spread out within each group. Every individual in the data set was grouped by their highest level of education, which allowed us to compare how education might relate to workload. Looking at these summary stats gave us a better understanding of how work hours vary depending on education and helped set the foundation for the visualizations and analysis we did later in the project. The summary table 2 is as below:

Table 2: Average Weekly Hours and Median Earnings by Major Group

Major_Group	Avg_Hours	Median_Earnings
Law & Public Policy	47.96	36000
Engineering	41.74	57000
Industrial Arts & Consumer Services	41.62	35000
Business	41.59	40000
Health	41.36	35000

Data Visualizations

After exploring our summary tables, we created several visualizations to better understand how work hours and earnings relate across different demographics and job fields. Each graph helped us dig deeper into our research question: "Does Working More Hours Result in Higher Pay*?* Below, we explain each visualization and what it helped us learn from our data.

Figure 1: Median Weekly Hours by Education Level

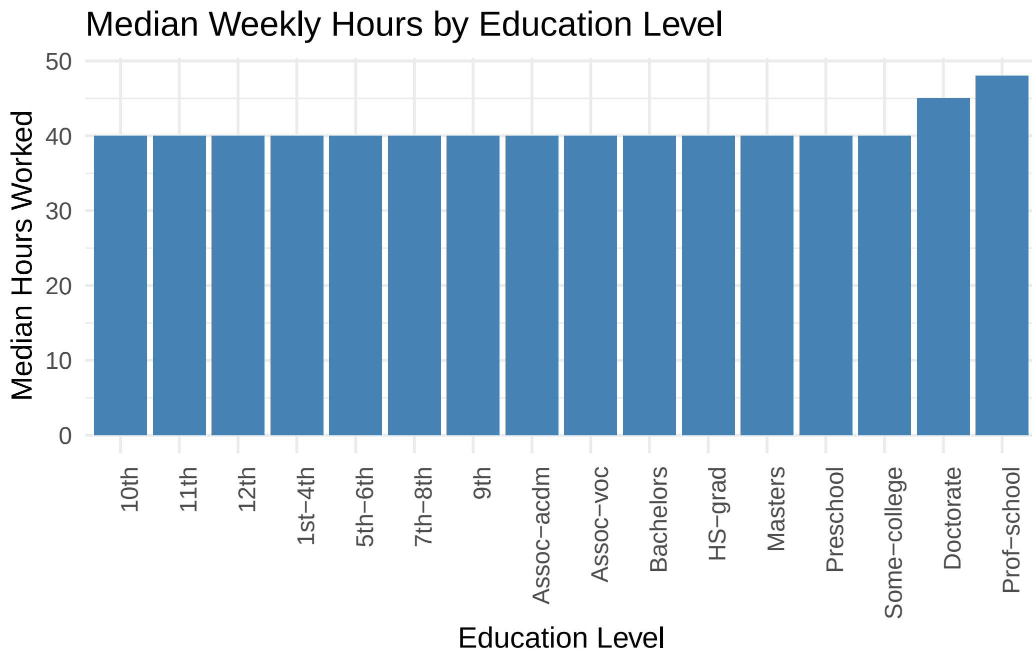
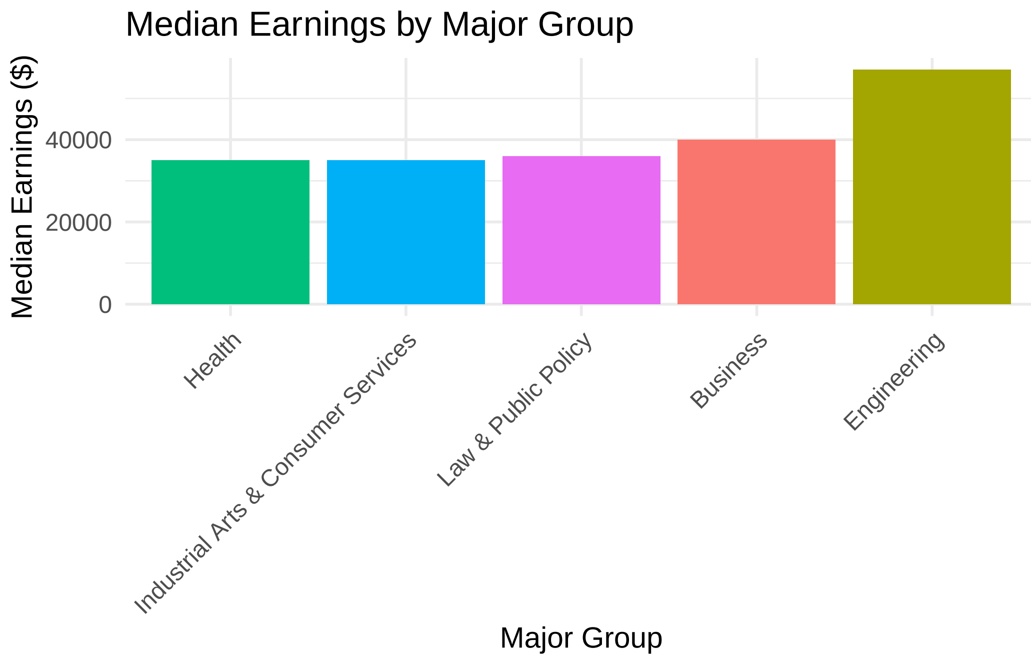


Figure 1 Our first bar chart displays the median number of hours worked each week across different education levels. We noticed that individuals with higher education, such as a Bachelor's or Master's degree, tend to work slightly more hours compared to those with lower levels of education. While this doesn't directly address pay, it helps us understand how work commitment might increase with educational attainment. It shows the typical number of hours worked weekly by education level. We observed that higher education levels, such as Bachelor's or Master's degrees, generally correlate with longer work hours, giving context for how education might impact work-life balance.

Figure 2: Median Earnings by Major Group



1. Next, we looked at the typical income by major group. Fields like Engineering and Business had the highest median earnings, while Education and Industrial Arts earned significantly less. This visualization helped show which degrees tend to pay off more financially, regardless of hours worked. It displays how earnings vary across fields of study. Business and Engineering majors had the highest median incomes, while fields like Education and Industrial Arts had lower. This helped us understand how major choice influences financial outcomes.

Figure 3: Average Weekly Hours by Major Group

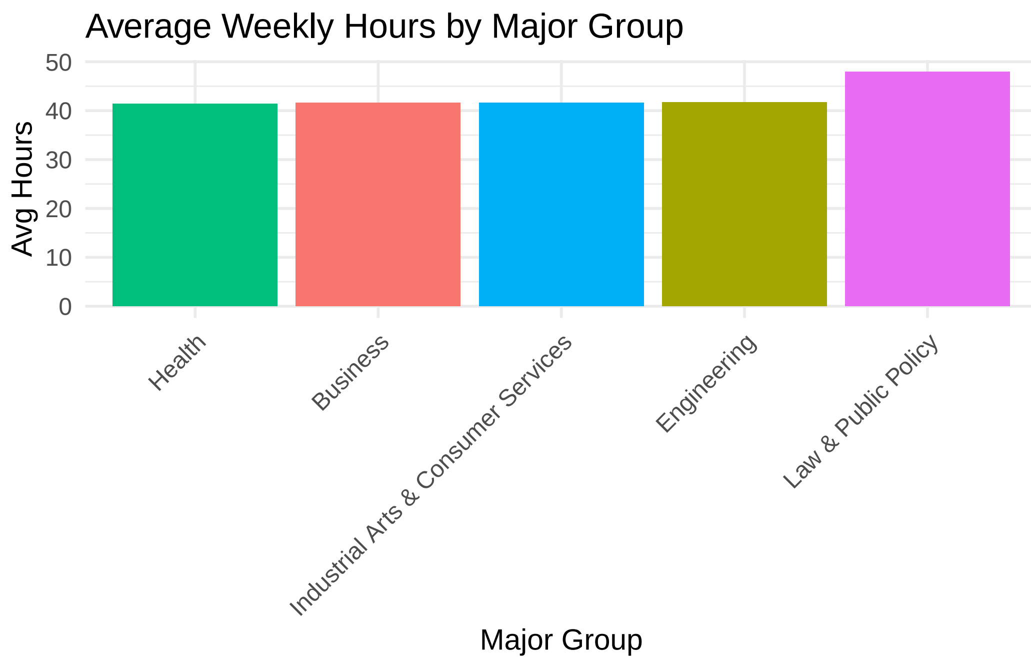


Figure 3 shows the average hours worked per week by major group. Interestingly, some fields like Law & Public Policy had high workloads but lower pay, showing that more hours don't always mean higher earnings. This bar chart shows how many hours people work on average in each major group. Interestingly, some groups like Law & Public Policy had the highest average weekly hours, yet their median earnings were lower than other fields that worked fewer hours. This was one of our first clear signs that working more hours doesn't always guarantee higher pay.

Figure 4: Hours Worked per Week by Sex

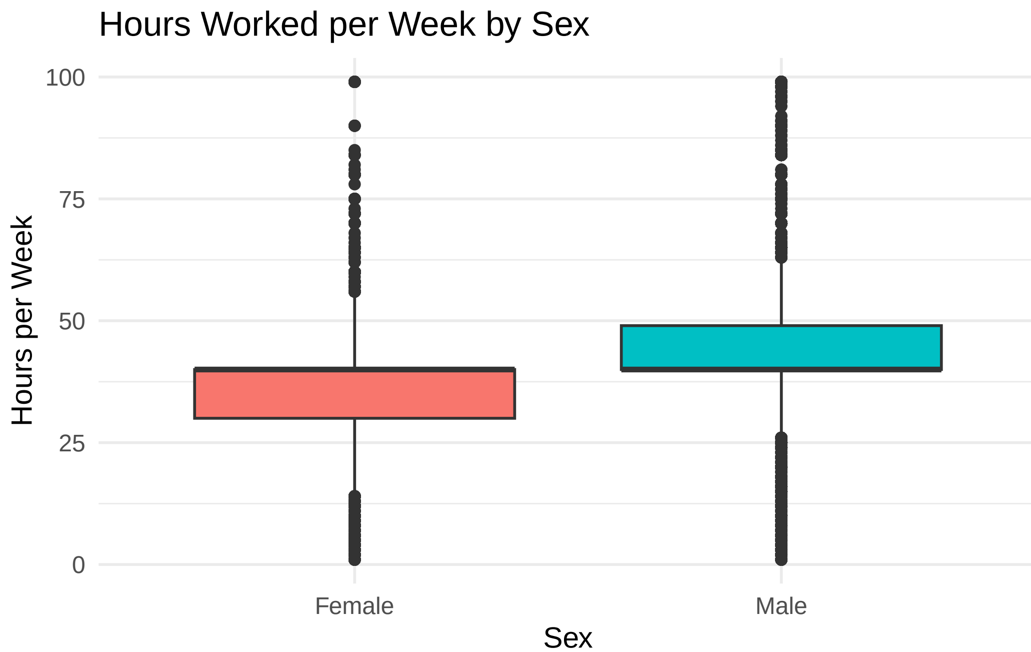


Figure 4 compares how many hours males and females work on average. We noticed males typically work more hours, which aligns with known workforce trends, although this doesn't account for income yet. This box plot compares the distribution of work hours between males and females. While there's an overlap, males tend to work slightly more hours per week on average. This supports broader labor trend data, though it doesn't tell us much about the income side yet.

Figure 5: Hours Worked by Education Level

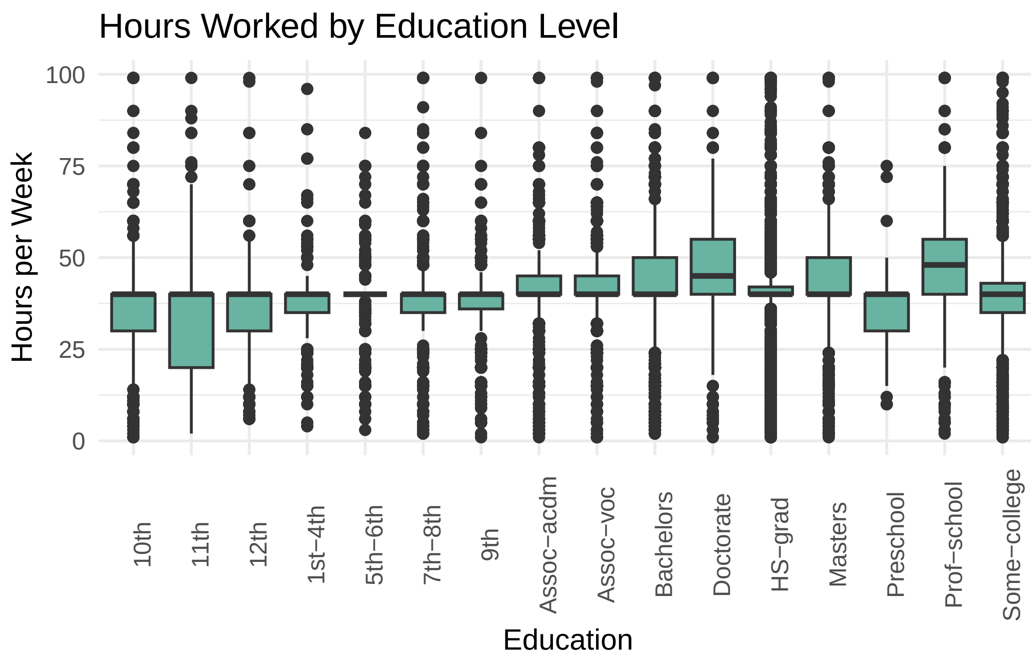


Figure 5 displays the spread of hours worked across education levels. Higher education is often linked with more hours, but the boxplot shows a wide range of outcomes even within each group. To go further, we used another box plot to see how work hours are spread within each education group. We found that while some higher education levels work more hours, the range within each group is wide — meaning not all college graduates work long hours, and not all people without degrees work fewer hours.

Figure 6: Hours Worked by Workclass

Figure 6 shows how work hours differ by employment type. Self-employed individuals show a wider range of hours, while private and government workers tend to have more consistent work schedules. This visualization breaks down work hours by employment type (private, self-employed, government, etc.). People in private and government roles tend to work more consistently, while those who are self-employed or without salary-based jobs show more variation in hours.

Figure 7: Heatmap of Average Weekly Hours by Age Group and Education

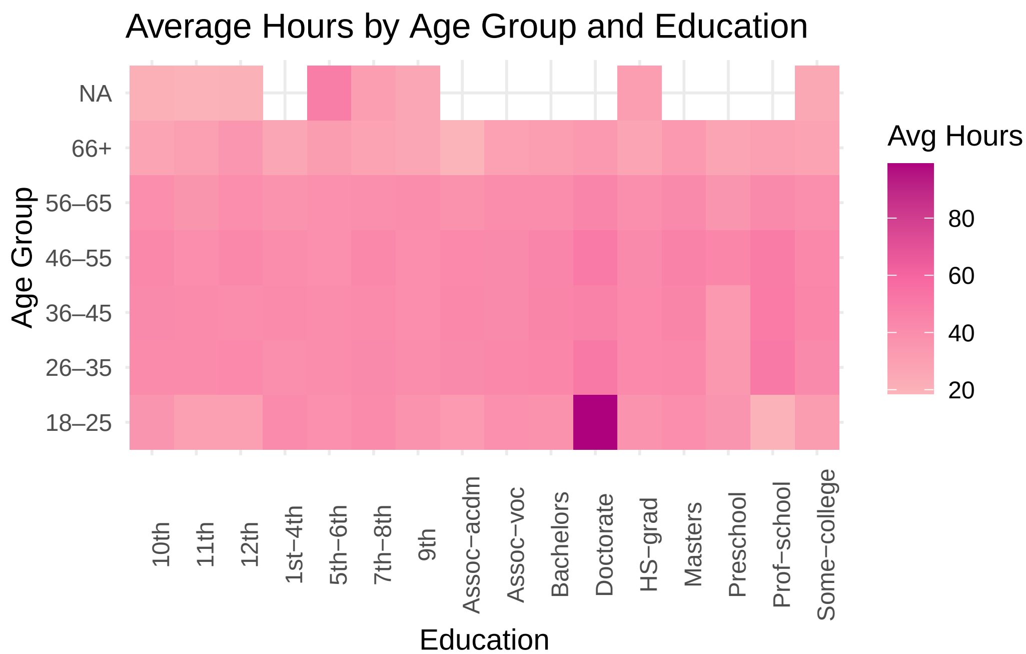
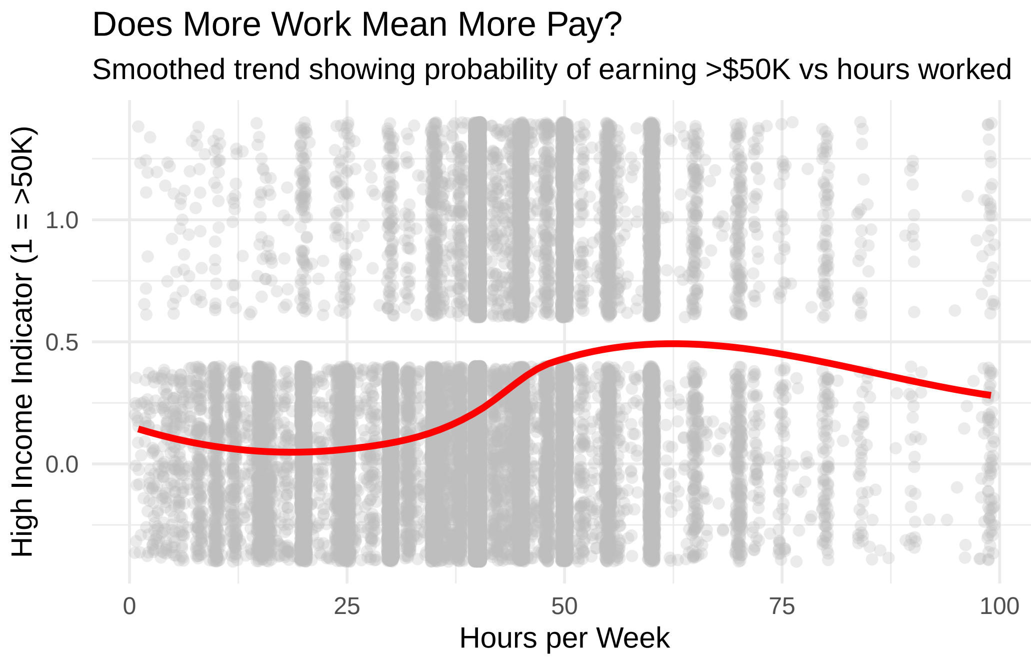


Figure 7 presents average weekly hours broken down by both age and education. Middle-aged adults tend to work the most, especially those with higher education levels, offering deeper insight into work patterns over time. We also created a heatmap that shows the average number of hours worked across different age groups and education levels. We found that middle-aged adults (36–55) typically work the most, and that trend holds across education levels. This gave us more context about where longer work weeks are most common.

Figure 8: Probability of Earning >$50K Based on Hours Worked



Lastly, Figure 8 directly answers our research question. The probability of earning more than $50K does increase with more hours worked, but flattens out around 50–60 hours per week, suggesting diminishing returns. We added a smoothed regression line to show the trend. The results showed that, up to a certain point, working more hours does increase the chance of earning more, but the curve flattens out after around 50–60 hours per week. This suggests diminishing returns, meaning working extreme hours doesn't always result in significantly higher income.

Conclusion

Through our analysis, we explored the relationship between hours worked, earnings, and various factors like education level, job type, and field of study. At first, we assumed that working more hours would always lead to higher pay, but our visualizations and summary tables revealed that this isn't always the case. While people in higher-paying fields like Business or Engineering often work full-time hours, we also found that some groups, such as those in Law & Public Policy, work longer hours but earn less. Our scatter plot confirmed that the probability of earning over $50K does increase with hours worked — but only to a point. After around 50–60 hours per week, the benefit starts to level off, suggesting that overworking doesn't always result in better pay.

Overall, this project helped us realize that income isn't only determined by how much you work, but also by what you do and your level of education. By breaking the data down into visuals and summaries, we were able to make sense of complex patterns and draw conclusions that can be helpful for students thinking about future careers. This project also reminded us of the importance of work-life balance — because more work doesn't always mean more money, and sometimes, it may not be worth the extra stress.

Reflection

Working on this project helped us realize how important it is to look at data instead of just assuming things. At first, we thought the answer to our research question would be obvious — that more hours means more money. But after analyzing the data, we found that it's a lot more complicated. It depends on what kind of job you have, what field you're in, and how much education you've completed.

We also learned a lot about using R and Quarto to turn messy datasets into clear visualizations and summaries. It was interesting to see how graphs and tables can make patterns easier to understand and help tell a story. This made us feel more confident working with real-world data and seeing how it connects to topics that matter, like career decisions and financial planning. Overall, this project made us more aware of the trade-offs between work and pay, and it reminded us that finding balance is just as important as chasing a bigger paycheck.

Work Cited

1. **Mulye, Aditi.** "Adult Income Dataset from Scratch." *Kaggle*, Nov. 2024, https://www.kaggle.com/code/aditimulye/adult-income-dataset-from-scratch.
2. **Hussain, Tayyar.** "Economic Values of College Majors: Data Analysis." *Kaggle*, Nov. 2024, https://www.kaggle.com/code/tayyarhussain/economic-values-of-college-majors-data-analysis.

Code Appendix

```
# Load libraries
library(tidyverse)

# Load raw data
adult_income <- read_csv("https://raw.githubusercontent.com/siyonabehera/Sec1_FP_NurAmeeraSabr
college_majors <- read_csv("https://raw.githubusercontent.com/siyonabehera/Sec1_FP_NurAmeeraSal

# Clean missing values
adult_income <- adult_income %>%
  mutate(across(where(is.character), ~na_if(., "?"))) %>%
  drop_na()

# Combine low education levels
adult_income <- adult_income %>%
  mutate(education = case_when(
    education %in% c("Preschool", "1st-4th", "5th-6th", "7th-8th",
                     "9th", "10th", "11th", "12th") ~ "Less than HS",
    TRUE ~ education
  ))
```

```
# Create Major_Group
adult_income <- adult_income %>%
  mutate(
    Major_Group = case_when(
      str_detect(education, "Bachelors|Some-college") & str_detect(occupation, "Engineer|Tech-s
      str_detect(education, "Bachelors|Some-college") & str_detect(occupation, "Exec-manageria
      str_detect(education, "Masters|Bachelors|Some-college") & str_detect(occupation, "Prof-s
      str_detect(education, "Prof-school") ~ "Law & Public Policy",
      str_detect(education, "Assoc-acdm|Assoc-voc") ~ "Industrial Arts & Consumer Services",
      TRUE ~ "Other"
    )
  )

# Keep relevant columns
adult_income <- adult_income %>%
  select(education, `hours-per-week`, sex, age, race, workclass, Major_Group)

# Clean college majors
college_majors <- college_majors %>%
  rename(Major_Group = Major_category) %>%
  select(Major, Major_Group, Median)

# View cleaned datasets
View(adult_income)
View(college_majors)

# Export (non-reproducible path warning)
write_csv(adult_income, "cleaned_adult_income.csv")
write_csv(college_majors, "cleaned_college_majors.csv")

## 2. Data Summarization and Merging

#| echo: true
#| eval: false

# Summary: median hours by education
education_hours_summary <- adult_income %>%
  group_by(education) %>%
  summarise(Median_Hours = median(`hours-per-week`, na.rm = TRUE)) %>%
  arrange(Median_Hours)

# Summary: average hours by major group
adult_summary <- adult_income %>%
  group_by(Major_Group) %>%
  summarise(Avg_Hours = mean(`hours-per-week`, na.rm = TRUE)) %>%
  arrange(desc(Avg_Hours))
```

```r
# Summary: median earnings by major group
college_summary <- college_majors %>%
  group_by(Major_Group) %>%
  summarise(Median_Earnings = median(Median, na.rm = TRUE)) %>%
  arrange(desc(Median_Earnings))

# Merge summaries
merged_summary <- inner_join(adult_summary, college_summary, by = "Major_Group")

# Check categories
unique(college_majors$Major_Group)
intersect(unique(adult_income$Major_Group), unique(college_majors$Major_Group))

# View tables
View(adult_summary)
View(college_summary)
View(merged_summary)

# Export merged summary
write_csv(merged_summary, "merged_summary.csv")


## 3. Visualizations and Plot Preparation

#| echo: true
#| eval: false

# Load cleaned/merged data
adult_income <- read_csv("https://raw.githubusercontent.com/siyonabehera/Sec1_FP_NurAmeeraSabri
merged_summary <- read_csv("https://raw.githubusercontent.com/siyonabehera/Sec1_FP_NurAmeeraSal

# Plot 1: Bar Chart - Median Hours by Education
ggplot(education_hours_summary, aes(x = reorder(education, Median_Hours), y = Median_Hours)) +
  geom_col(fill = "steelblue") +
  labs(
    title = "Median Weekly Hours by Education Level",
    subtitle = "Ordered by increasing hours",
    x = "Education Level",
    y = "Median Hours Worked"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

# Plot 2: Bar Chart - Median Earnings by Major Group
ggplot(merged_summary, aes(x = reorder(Major_Group, Median_Earnings), y = Median_Earnings, fil
  geom_col() +
  labs(title = "Median Earnings by Major Group", x = "Major Group", y = "Median Earnings ($)")
```

```
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill = "none")

# Plot 3: Bar Chart - Avg Weekly Hours by Major Group
ggplot(merged_summary, aes(x = reorder(Major_Group, Avg_Hours), y = Avg_Hours, fill = Major_Gro
  geom_col() +
  labs(title =  "Hours Worked per Week by Major Group on Average", x = "Major Group", y = "Aver
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill = "none")

# Plot 4: Boxplot - Hours per Week by Sex
ggplot(adult_income, aes(x = sex, y = `hours-per-week`, fill = sex)) +
  geom_boxplot() +
  labs(title = "Hours Worked per Week by Sex", x = "Sex", y = "Hours per Week") +
  theme_minimal() +
  guides(fill = "none")

# Plot 5: Boxplot - Hours per Week by Education
ggplot(adult_income, aes(x = education, y = `hours-per-week`)) +
  geom_boxplot(fill = "#69b3a2") +
  labs(title = "Hours Worked per Week by Education Level", x = "Education", y = "Hours per Weel
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90))

# Plot 6: Boxplot - Hours per Week by Workclass
ggplot(adult_income, aes(x = workclass, y = `hours-per-week`, fill = workclass)) +
  geom_boxplot() +
  labs(title = "Hours Worked per Week by Workclass", x = "Workclass", y = "Hours per Week") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill = "none")

# Add Age_Group + income_binary
adult_income <- adult_income %>%
  mutate(
    Age_Group = cut(age, breaks = c(17, 25, 35, 45, 55, 65, 90),
                    labels = c("18-25", "26-35", "36-45", "46-55", "56-65", "66+")),
    income_binary = if_else(income == ">50K", 1, 0)
  )

# Heatmap data
heatmap_data <- adult_income %>%
  group_by(Age_Group, education) %>%
  summarise(Avg_Hours = mean(`hours-per-week`, na.rm = TRUE))
```

```r
# Plot 7: Heatmap - Avg Hours by Age Group and Education
ggplot(heatmap_data, aes(x = education, y = Age_Group, fill = Avg_Hours)) +
  geom_tile() +
  labs(title = "Average Hours by Age Group and Education", x = "Education", y = "Age Group", fi
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90)) +
  scale_fill_gradientn(colors = c("#fbb4b9", "#f768a1", "#ae017e"))

# Plot 8: Scatter - Hours vs High Income
ggplot(adult_income, aes(x = `hours-per-week`, y = income_binary)) +
  geom_jitter(alpha = 0.3, color = "gray") +
  geom_smooth(method = "loess", se = FALSE, color = "red", size = 1.2) +
  labs(title = "Does More Work Mean More Pay?",
       subtitle = "Smoothed trend showing probability of earning >$50K vs hours worked",
       x = "Hours per Week",
       y = "High Income Indicator (1 = >50K)") +
  theme_minimal()
```