Artificial Intelligence Opinion Survey

DATA 490 Independent Study

Mentor: Dr. Nicholas Dietrich
Aadarsha Gopala Reddy, Darren Lo,
Ethan Love, Jose Mancilla, Santo Sumo

April 20, 2023

Contents

1. Load Data

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.1     v purrr   1.0.1
## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.4     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)

# Load data. Top row is column name.
edu <- read.csv("prolific_edu.csv")
health <- read.csv("prolific_health.csv")
retail <- read.csv("prolific_retail.csv")
tech <- read.csv("prolific_tech.csv")
qualtrics <- read.csv("qualtrics_data.csv")
```

2. Data Cleaning

```
# Combine data into one data frame after mutating Age to be one data type
edu <- edu %>% mutate(Age = as.character(Age))
```

```r
health <- health %>% mutate(Age = as.character(Age))
retail <- retail %>% mutate(Age = as.character(Age))
tech <- tech %>% mutate(Age = as.character(Age))
combined <- bind_rows(edu, health, retail, tech)
# export combined data to csv
# write.csv(combined, "combined_non_qualtrics.csv")
```

```r
# combine qualtrics and combined data using qualtrics data's
#     ProlificID column and combined data's Participant id
combined <- left_join(qualtrics, combined, by = c("ProlificID" = "Participant.id"))
# rename "Duration..in.seconds." column to "Duration"
colnames(combined)[colnames(combined) == "Duration..in.seconds."] <- "Duration"
# remove Age.x and keep only Age.y column and rename Age.y to Age
combined <- combined %>%
  select(-Age.x) %>%
  rename(Age = Age.y)
# remove Status.x and Status.y columns
combined <- combined %>%
  select(-Status.x) %>%
  select(-Status.y)
# remove Finished, Progress, UserLanguage, DistributionChannel,
# Nationality, and Consent columns
combined <- combined %>%
  select(-Finished) %>%
  select(-Progress) %>%
  select(-UserLanguage) %>%
  select(-DistributionChannel) %>%
  select(-Nationality) %>%
  select(-Consent)

# remove rows where Submission.id is NA
combined <- combined %>% filter(!is.na(Submission.id))
# Keep only rows which say "United States" in "Country.of.residence" column
combined <- combined %>% filter(combined$Country.of.residence == "United States")

# export data to csv
# write.csv(combined, "combined_qualtrics.csv")
```

```r
# Keep only rows which say "Compose an email" in "Attention" column
combined <- combined %>% filter(combined$Attention == "Compose an email")
# remove Attention column
combined <- combined %>% select(-Attention)
# export data to csv
# write.csv(combined, "combined_qualtrics_attentive.csv")
```

3. Data Exploration

The columns in the dataset are:

- StartDate - Date and time survey was started
- EndDate - Date and time survey was completed
- IPAddress - IP address of participant

- Duration - Duration of survey in seconds
- RecordedDate - Date and time survey was recorded
- ResponseId - Response ID
- LocationLatitude - Participant's location latitude
- LocationLongitude - Participant's location longitude
- ProlificID - Identification of the response on Prolific
- Gender - Gender of the participant
- Education - Education level of the participant
- Salary - Salary of the participant
- AIKnowledge - Knowledge of AI of the participant
- UsedAI - Whether the participant has used AI
- TimeEnergy - How much time and energy AI has saved the participant
- SimilarTasks - How much of the participant's tasks they think AI can do
- EnhanceHurt - Whether the participant thinks AI can enhance or hurt their work efficiency.
- Comments - Comments from the participant
- Submission.id - Submission ID
- Started.at - Date and time survey was started
- Completed.at - Date and time survey was completed
- Reviewed.at - Date and time survey was reviewed
- Archived.at - Date and time survey was archived
- Time.taken - Duration of survey in seconds
- Completion.code - Completion code
- Total.approvals - Total number of approvals
- Employment.sector - Employment sector
- Age - Age of the participant
- Sex - Sex of the participant
- Ethnicity.simplified - Ethnicity of the participant
- Country.of.birth - Country of birth of the participant
- Country.of.residence - Country of residence of the participant
- Language - Language of the participant
- Student.status - Whether the participant is a student
- Employment.status - Whether the participant is employed

4. Data Analysis

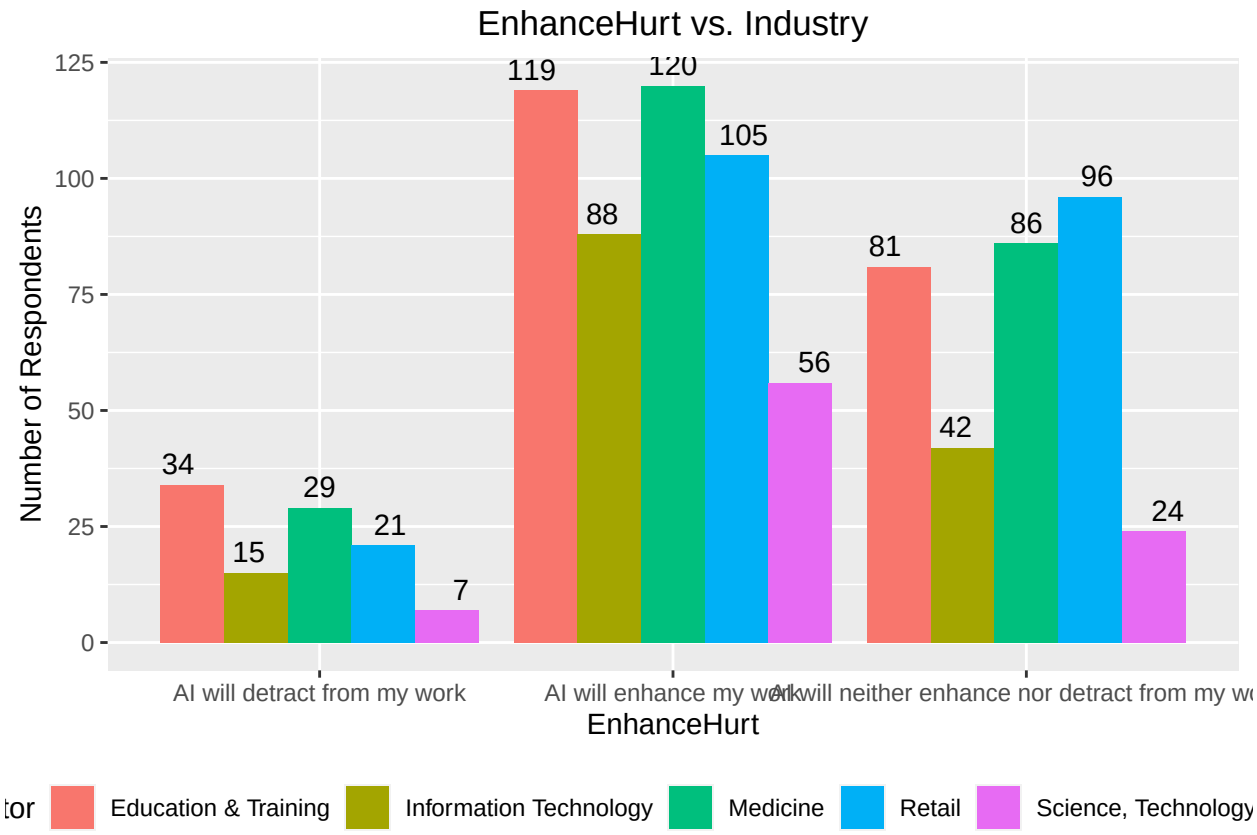
4.1. EnhanceHurt vs. Industry

```
# Create a new data frame with only the columns we need
enhancehurt_vs_industry <- combined %>%
  select(EnhanceHurt, Employment.sector)

colnames(enhancehurt_vs_industry)
```

```
## [1] "EnhanceHurt"       "Employment.sector"
```

```
# Remove rows where Employment.sector is NA
enhancehurt_vs_industry <- enhancehurt_vs_industry %>%
  filter(!is.na(Employment.sector))
# Remove rows where EnhanceHurt is NA
enhancehurt_vs_industry <- enhancehurt_vs_industry %>%
  filter(!is.na(EnhanceHurt))

# Visualize using different histogram for each industry.
# Show the number of respondents on each bar for each choice.
ggplot(enhancehurt_vs_industry, aes(x = EnhanceHurt, fill = Employment.sector)) +
  geom_bar(position = "dodge") +
  geom_text(stat = "count", aes(label = ..count..), position = position_dodge(width = 1), vjust = -0.5)
  labs(title = "EnhanceHurt vs. Industry", x = "EnhanceHurt", y = "Number of Respondents") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "bottom")
```

```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(count)` instead.
```


EnhanceHurt vs. Industry
Number of Respondents
125
100
75
50
25
0
34
15
29
21
7
119
88
120
105
56
81
42
86
96
24
AI will detract from my work
AI will enhance my work
AI will neither enhance nor detract from my w
EnhanceHurt
tor
Education & Training
Information Technology
Medicine
Retail
Science, Technology