

Reproducible Research Project: Technical and Humanities Students' Perspectives on the Development and Sustainability of Artificial Intelligence (AI)

Szymon Karkoszka, Antoni Piotrowski

Loading required packages

```
library(rstudioapi)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2     3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
```

```
## -- 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 errors
```

```
library(RColorBrewer)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##      smiths
```

```
library(ggplot2)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##      group_rows
```

1. Why it is important to reproduce the paper — problem description and topic importance

The perception of students on AI in education is a crucial topic to analyze from an economic standpoint. The incorporation of AI in the education sector has been increasing in recent years, with the emergence of smart classrooms and personalized learning platforms. The potential benefits of AI in education include improved learning outcomes, increased accessibility, and efficiency gains in educational delivery. Understanding the perceptions of students towards AI in education is vital to inform policies and investment decisions in the

education sector. The analysis of student perceptions can provide insights into the potential barriers and facilitators to AI adoption in education, as well as inform strategies for the development of AI technologies that are sensitive to student needs and concerns.

This paper aims to analyze students' perception on the AI in Education. Therefore we decide to examine the differences between the studies.

2. Data In terms of the comparison, the data we found are excellent as they both compare student and both from Romania. Unfortunately, our data set is *10 times smaller* but much more informative. Our data are also much more up to date and we can say that we compare data between two “AI Waves” — one from 2017 and another from 2022.

The dataset utilized in this study encompasses the responses obtained from a survey administered to undergraduate students in their 2nd and 3rd years of study at the Faculty of Cybernetics, Statistics, and Economic Informatics. The survey was conducted online and disseminated through social media groups. Its objective was to acquire insights into students' perceptions regarding the role of artificial intelligence (AI) in the domain of education.

This dataset comprises a total of 96 observations, each representing a unique case or participant. Additionally, the dataset contains 36 variables, which encompass the different characteristics or attributes under investigation.

Loading the data

```
current_path <- rstudioapi::getActiveDocumentContext()$path
current_path <- path.expand(current_path)
current_path <- dirname(current_path)
to_csv <- file.path(current_path, "Data", "Survey_AI.csv")
to_csv
```

```
## [1] "/Users/szymonkarkoszka/Desktop/RR_FinalProject/RR_fp/Data/Survey_AI.csv"
```

```
data <- read.csv(to_csv)
```

Checking the structure of the data

```
str(data)
```

```
## 'data.frame':   91 obs. of  35 variables:
## $ ID : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Q1.AI_knowledge : int  8 7 5 5 4 5 7 6 6 4 ...
## $ Q2.AI_sources : chr  "Internet;Books/Scientific papers (physical/online form
## $ Q2.1.Internet : int  1 1 1 1 1 1 0 1 1 1 ...
## $ Q2.2.Books.Papers : int  1 0 1 0 0 0 1 0 1 0 ...
## $ Q2.3.Social_media : int  0 1 0 1 0 0 1 0 0 0 ...
## $ Q2.4.Discussions : int  0 0 0 0 0 1 1 0 1 0 ...
## $ Q2.5.NotInformed : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Q3.1.AI_dehumanization : int  1 2 2 4 1 3 1 1 2 4 ...
## $ Q3.2.Job_replacement : int  2 3 1 4 2 4 2 1 4 4 ...
## $ Q3.3.Problem_solving : int  5 4 4 5 5 4 4 5 5 4 ...
## $ Q3.4.AI_rulling_society : int  1 1 1 3 1 3 1 1 2 3 ...
## $ Q4.1.AI_costly : int  4 3 3 4 3 4 5 3 5 4 ...
## $ Q4.2.Economic_crisis : int  2 3 1 3 1 3 4 2 2 3 ...
## $ Q4.3.Economic_growth : int  4 4 3 3 4 3 2 3 3 3 ...
## $ Q4.4.Job_loss : int  2 3 2 4 2 4 4 2 4 4 ...
## $ Q5.Feelings : int  1 1 1 1 1 2 1 1 1 1 ...
## $ Q6.Domains : chr  "Education;Medicine;Marketing" "Medicine;Agriculture;C
## $ Q6.1.Education : int  1 0 1 1 1 1 1 1 0 1 ...
## $ Q6.2.Medicine : int  1 1 0 1 1 0 1 1 0 1 ...
```

```
## $ Q6.3.Agriculture      : int  0 1 0 0 1 1 0 0 1 1 ...
## $ Q6.4.Constructions    : int  0 1 0 0 1 0 1 0 0 1 ...
## $ Q6.5.Marketing        : int  1 0 1 0 0 0 0 0 1 0 ...
## $ Q6.6.Administration   : int  0 0 1 0 0 1 0 0 1 0 ...
## $ Q6.7.Art              : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Q7.Utility_grade      : int  9 6 6 9 8 6 10 8 8 7 ...
## $ Q8.Advantage_teaching : int  3 2 3 1 3 1 1 3 3 1 ...
## $ Q9.Advantage_learning : int  1 2 3 2 2 2 2 3 2 1 ...
## $ Q10.Advantage_evaluation : int  2 1 3 2 3 3 1 3 2 3 ...
## $ Q11.Disadvantage_educational_process: int  3 2 4 3 4 1 3 3 1 1 ...
## $ Q12.Gender            : int  1 2 2 1 1 1 1 2 1 1 ...
## $ Q13.Year_of_study     : int  2 2 2 2 2 2 2 2 2 2 ...
## $ Q14.Major             : int  2 2 2 2 2 2 2 2 2 3 ...
## $ Q15.Passed_exams      : int  1 1 0 1 1 1 0 0 0 1 ...
## $ Q16.GPA               : num  9.2 7.7 7.2 8.2 7.7 7.7 7.2 6.7 7.7 8.2 ...
```

Deleting character variables (they're only as comments for the variables for Q2.1-Q2.5 and Q6.1-Q6.7)

```
data <- subset(data, select = -c(Q2.AI_sources, Q6.Domains, ID))
```

Changing all variables to numeric

```
data <- mutate_all(data, as.numeric)
```

Checking for missing values

```
print(missing_values <- colSums(is.na(data)))
```

```
##           Q1.AI_knowledge           Q2.1.Internet
##           0                      0
##           Q2.2.Books.Papers       Q2.3.Social_media
##           0                      0
##           Q2.4.Discussions         Q2.5.NotInformed
##           0                      0
##           Q3.1.AI_dehumanization   Q3.2.Job_replacement
##           0                      0
##           Q3.3.Problem_solving      Q3.4.AI_rulling_society
##           0                      0
##           Q4.1.AI_costly           Q4.2.Economic_crisis
##           0                      0
##           Q4.3.Economic_growth      Q4.4.Job_loss
##           0                      0
##           Q5.Feelings              Q6.1.Education
##           0                      0
##           Q6.2.Medicine            Q6.3.Agriculture
##           0                      0
##           Q6.4.Constructions        Q6.5.Marketing
##           0                      0
##           Q6.6.Administration       Q6.7.Art
##           0                      0
##           Q7.Utility_grade          Q8.Advantage_teaching
##           0                      0
##           Q9.Advantage_learning     Q10.Advantage_evaluation
##           0                      0
##           Q11.Disadvantage_educational_process Q12.Gender
##           0                      0
##           Q13.Year_of_study         Q14.Major
```

```
##                                0                                0
##                                Q15.Passed_exams                    Q16.GPA
##                                0                                0
```

Creating data frame with sources of students' knowledge with their counts and producing an appropriate plot

Comparison with the paper: A first perspective they identify how well the respondents understood the concept of artificial intelligence. Their results show that 84.6% consider that they know what artificial intelligence means, whereas 12% claim that they do not know what the concept means. At the same time, 3.4% answered "I don't know/I won't answer"

The highest percentage belongs to the category of those who believe that they have an average awareness of artificial intelligence (36.1%). By referring to this average category, while cumulating two other answers (very well informed, 2.4%, and well informed, 16.8%), which represent the number of those with an above-average level of awareness, as well as cumulating the remaining two answers (poorly informed, 35%, and very poorly informed, 9.7%) which represent the category of those with a below-average level of awareness, showing that the number of those who consider themselves less informed is greater than the category of those who believe that they are well informed on the subject (44.7% as opposed to 19.2%).

// weź wrzuc tutaj jakiś plot w porównaniu z naszymi danymi.

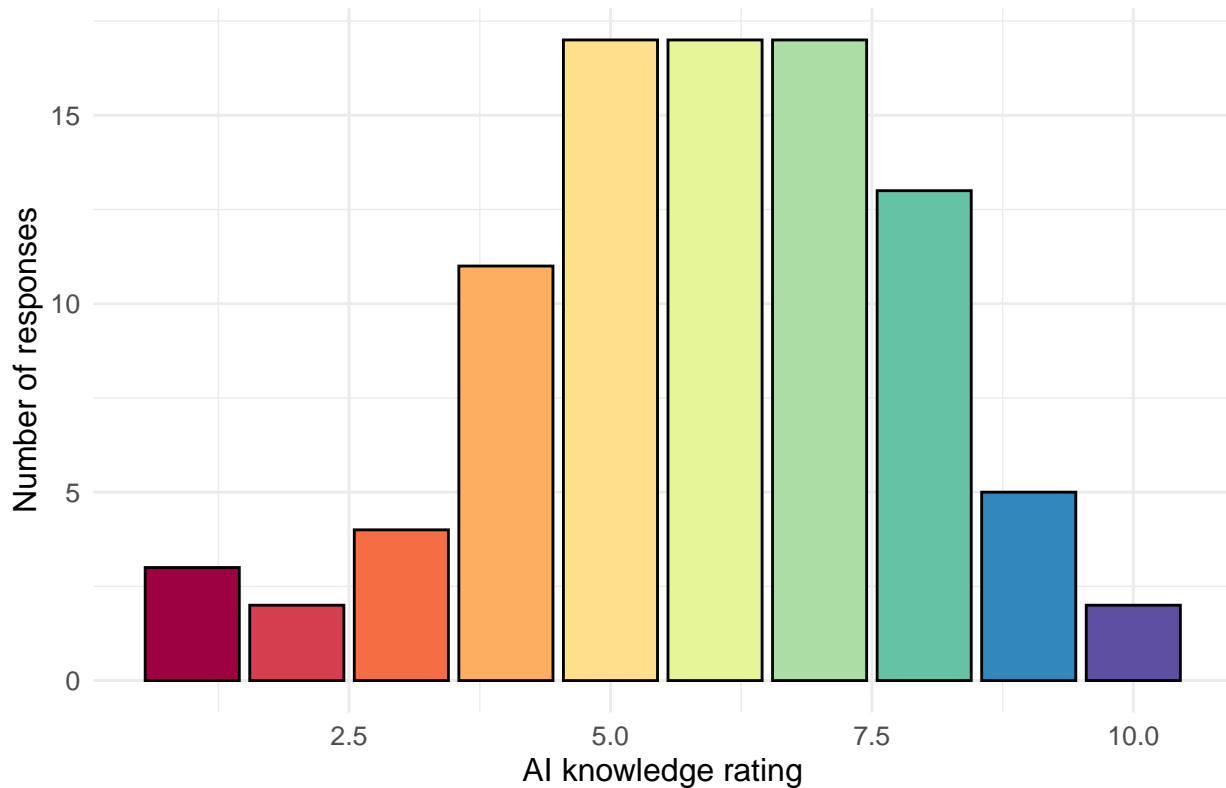
No missing values found, so we can proceed with the barplot of students' knowledge

```
custom_colors <- brewer.pal(10, "Spectral")

fig1 <- ggplot(data, aes(x = Q1.AI_knowledge)) +
  geom_bar(aes(fill = factor(Q1.AI_knowledge)), color = "black") +
  scale_fill_manual(values = custom_colors) +
  xlab("AI knowledge rating") +
  ylab("Number of responses") +
  ggtitle("A countplot of AI knowledge ratings") +
  theme_minimal() +
  theme(plot.title = element_text(size = 16, face = "bold"),
        axis.title = element_text(size = 12),
        axis.text = element_text(size = 10),
        legend.title = element_blank(),
        legend.position = "none")

print(fig1)
```

A countplot of AI knowledge ratings



```
source_names <- c("Internet", "Books/Papers", "Social Media", "Discussions", "Not Informed")
source_counts <- sapply(data[, c("Q2.1.Internet", "Q2.2.Books.Papers", "Q2.3.Social_media", "Q2.4.Discu

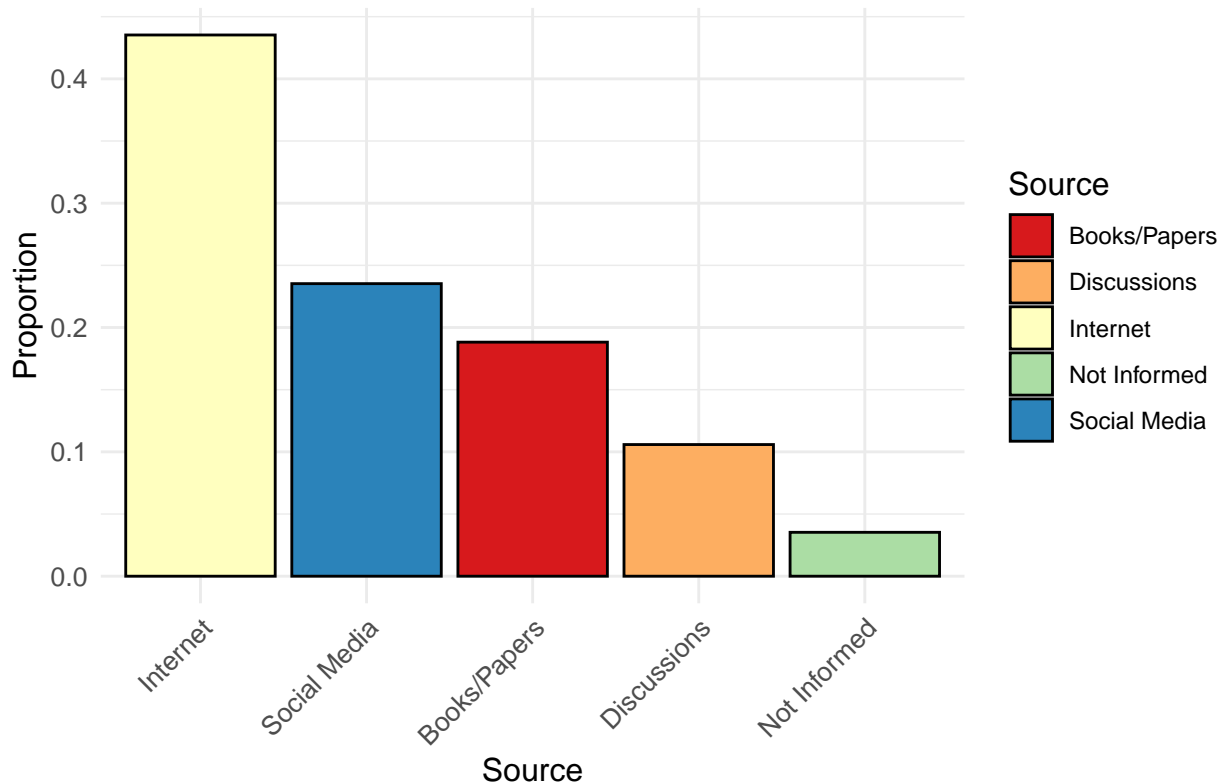
source_data <- data.frame(Source = source_names, Count = source_counts)

source_data$Proportion <- source_data$Count / sum(source_data$Count)

custom_colors <- brewer.pal(5, "Spectral")
fig2 <- ggplot(source_data, aes(x = reorder(Source, -Proportion), y = Proportion, fill = Source)) +
  geom_bar(stat = "identity", color = "black") +
  labs(x = "Source", y = "Proportion", fill = "Source") +
  ggtitle("Proportion of Students Using Each Source") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(size = 16, face = "bold"),
        axis.title = element_text(size = 12),
        axis.text = element_text(size = 10),
        legend.title = element_text(size = 12),
        legend.position = "right") +
  scale_fill_manual(values = custom_colors)

print(fig2)
```

Proportion of Students Using Each Source



correlations

Select the relevant columns from the data frame

```
selected_columns <- c("Q2.1.Internet", "Q2.2.Books.Papers", "Q2.3.Social_media", "Q2.4.Discussions", "Q2.5.NotInformed")
selected_data <- data[, selected_columns]
```

Calculate the correlation matrix and round the values

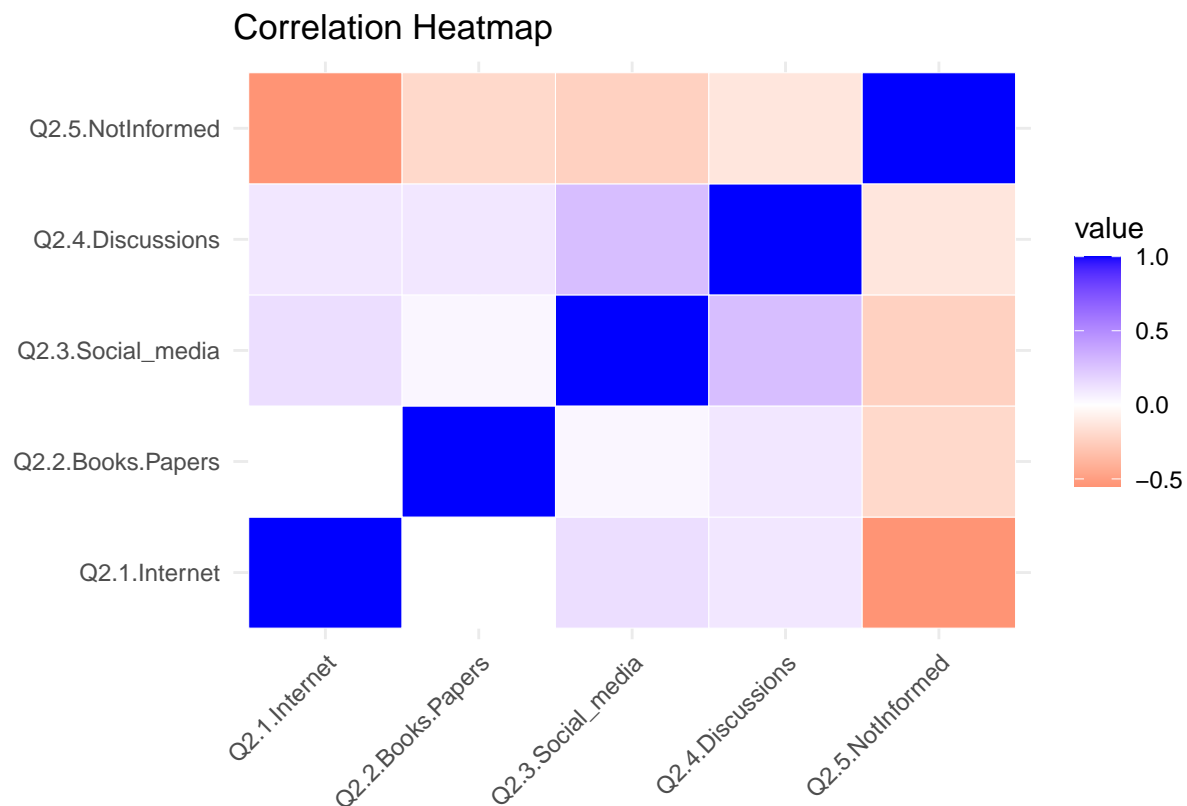
```
correlation_matrix <- round(cor(selected_data), 2)
correlation_data <- melt(correlation_matrix)
```

```
print(correlation_data)
```

```
##           Var1           Var2 value
## 1  Q2.1.Internet Q2.1.Internet  1.00
## 2  Q2.2.Books.Papers Q2.1.Internet  0.00
## 3  Q2.3.Social_media Q2.1.Internet  0.14
## 4  Q2.4.Discussions Q2.1.Internet  0.10
## 5  Q2.5.NotInformed Q2.1.Internet -0.55
## 6  Q2.1.Internet Q2.2.Books.Papers  0.00
## 7  Q2.2.Books.Papers Q2.2.Books.Papers  1.00
## 8  Q2.3.Social_media Q2.2.Books.Papers  0.04
## 9  Q2.4.Discussions Q2.2.Books.Papers  0.10
## 10 Q2.5.NotInformed Q2.2.Books.Papers -0.20
## 11 Q2.1.Internet Q2.3.Social_media  0.14
```

```
## 12 Q2.2.Books.Papers Q2.3.Social_media 0.04
## 13 Q2.3.Social_media Q2.3.Social_media 1.00
## 14 Q2.4.Discussions Q2.3.Social_media 0.28
## 15 Q2.5.NotInformed Q2.3.Social_media -0.24
## 16 Q2.1.Internet Q2.4.Discussions 0.10
## 17 Q2.2.Books.Papers Q2.4.Discussions 0.10
## 18 Q2.3.Social_media Q2.4.Discussions 0.28
## 19 Q2.4.Discussions Q2.4.Discussions 1.00
## 20 Q2.5.NotInformed Q2.4.Discussions -0.13
## 21 Q2.1.Internet Q2.5.NotInformed -0.55
## 22 Q2.2.Books.Papers Q2.5.NotInformed -0.20
## 23 Q2.3.Social_media Q2.5.NotInformed -0.24
## 24 Q2.4.Discussions Q2.5.NotInformed -0.13
## 25 Q2.5.NotInformed Q2.5.NotInformed 1.00
```

```
ggplot(data = correlation_data, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "red", mid = "white", high = "blue", midpoint = 0) +
  theme_minimal() +
  labs(title = "Correlation Heatmap", x = "", y = "") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```



Relation with student's knowledge about AI

```
# Define the columns vector
columns <- c("Q2.1.Internet", "Q2.2.Books.Papers", "Q2.3.Social_media", "Q2.4.Discussions", "Q2.5.NotInformed")

# Create an empty list to store the results
results <- list()
```

```

# Iterate over the columns
for (i in columns) {
  x1 <- mean(data[data[, i] == 0, "Q1.AI_knowledge"], na.rm = TRUE)
  x2 <- mean(data[data[, i] == 1, "Q1.AI_knowledge"], na.rm = TRUE)
  change <- round(x2 - x1, 2)
  results[[i]] <- c(x1, x2, change)
}

# Create a data frame from the results list
data2 <- data.frame(t(as.data.frame(results)))
colnames(data2) <- c("No - MeanQ1", "Yes - MeanQ1", "Absolute Change")
row.names(data2) <- columns

# Print the resulting data frame
print(data2)

```

```

##           No - MeanQ1 Yes - MeanQ1 Absolute Change
## Q2.1.Internet      5.000000    6.121622          1.12
## Q2.2.Books.Papers  5.423729    6.812500          1.39
## Q2.3.Social_media  5.411765    6.550000          1.14
## Q2.4.Discussions   5.849315    6.166667          0.32
## Q2.5.NotInformed   6.129412    2.833333         -3.30

```

```

# Define the column names and labels
columns <- c('Q3.1.AI_dehumanization', 'Q3.2.Job_replacement', 'Q3.3.Problem_solving', 'Q3.4.AI_rulling_society',
             'Q4.1.AI_costly', 'Q4.2.Economic_crisis', 'Q4.3.Economic_growth', 'Q4.4.Job_loss')
labels <- c('AI encourages dehumanization', 'Robots will replace people at work', 'AI helps to solve many problems',
            'AI will rule society', 'Machinery using AI is very expensive and resource-intensive to build',
            'AI will lead to a global economic crisis', 'AI will help global economic growth', 'AI leads to a better future')

# Subset the dataframe with the relevant columns
data_subset <- data[, columns]

str(data)

```

```

## 'data.frame':   91 obs. of  32 variables:
## $ Q1.AI_knowledge      : num  8 7 5 5 4 5 7 6 6 4 ...
## $ Q2.1.Internet        : num  1 1 1 1 1 1 0 1 1 1 ...
## $ Q2.2.Books.Papers    : num  1 0 1 0 0 0 1 0 1 0 ...
## $ Q2.3.Social_media     : num  0 1 0 1 0 0 1 0 0 0 ...
## $ Q2.4.Discussions      : num  0 0 0 0 0 1 1 0 1 0 ...
## $ Q2.5.NotInformed     : num  0 0 0 0 0 0 0 0 0 0 ...
## $ Q3.1.AI_dehumanization : num  1 2 2 4 1 3 1 1 2 4 ...
## $ Q3.2.Job_replacement  : num  2 3 1 4 2 4 2 1 4 4 ...
## $ Q3.3.Problem_solving  : num  5 4 4 5 5 4 4 5 5 4 ...
## $ Q3.4.AI_rulling_society : num  1 1 1 3 1 3 1 1 2 3 ...
## $ Q4.1.AI_costly        : num  4 3 3 4 3 4 5 3 5 4 ...
## $ Q4.2.Economic_crisis  : num  2 3 1 3 1 3 4 2 2 3 ...
## $ Q4.3.Economic_growth  : num  4 4 3 3 4 3 2 3 3 3 ...
## $ Q4.4.Job_loss         : num  2 3 2 4 2 4 4 2 4 4 ...
## $ Q5.Feelings           : num  1 1 1 1 1 2 1 1 1 1 ...
## $ Q6.1.Education        : num  1 0 1 1 1 1 1 1 0 1 ...
## $ Q6.2.Medicine         : num  1 1 0 1 1 0 1 1 0 1 ...
## $ Q6.3.Agriculture      : num  0 1 0 0 1 1 0 0 1 1 ...
## $ Q6.4.Constructions    : num  0 1 0 0 1 0 1 0 0 1 ...

```



```
## $ Q6.5.Marketing : num 1 0 1 0 0 0 0 0 1 0 ...
## $ Q6.6.Administration : num 0 0 1 0 0 1 0 0 1 0 ...
## $ Q6.7.Art : num 0 0 0 0 0 0 0 0 0 0 ...
## $ Q7.Utility_grade : num 9 6 6 9 8 6 10 8 8 7 ...
## $ Q8.Advantage_teaching : num 3 2 3 1 3 1 1 3 3 1 ...
## $ Q9.Advantage_learning : num 1 2 3 2 2 2 2 3 2 1 ...
## $ Q10.Advantage_evaluation : num 2 1 3 2 3 3 1 3 2 3 ...
## $ Q11.Disadvantage_educational_process: num 3 2 4 3 4 1 3 3 1 1 ...
## $ Q12.Gender : num 1 2 2 1 1 1 1 2 1 1 ...
## $ Q13.Year_of_study : num 2 2 2 2 2 2 2 2 2 2 ...
## $ Q14.Major : num 2 2 2 2 2 2 2 2 2 3 ...
## $ Q15.Passed_exams : num 1 1 0 1 1 1 0 0 0 1 ...
## $ Q16.GPA : num 9.2 7.7 7.2 8.2 7.7 7.7 7.2 6.7 7.7 8.2 ...

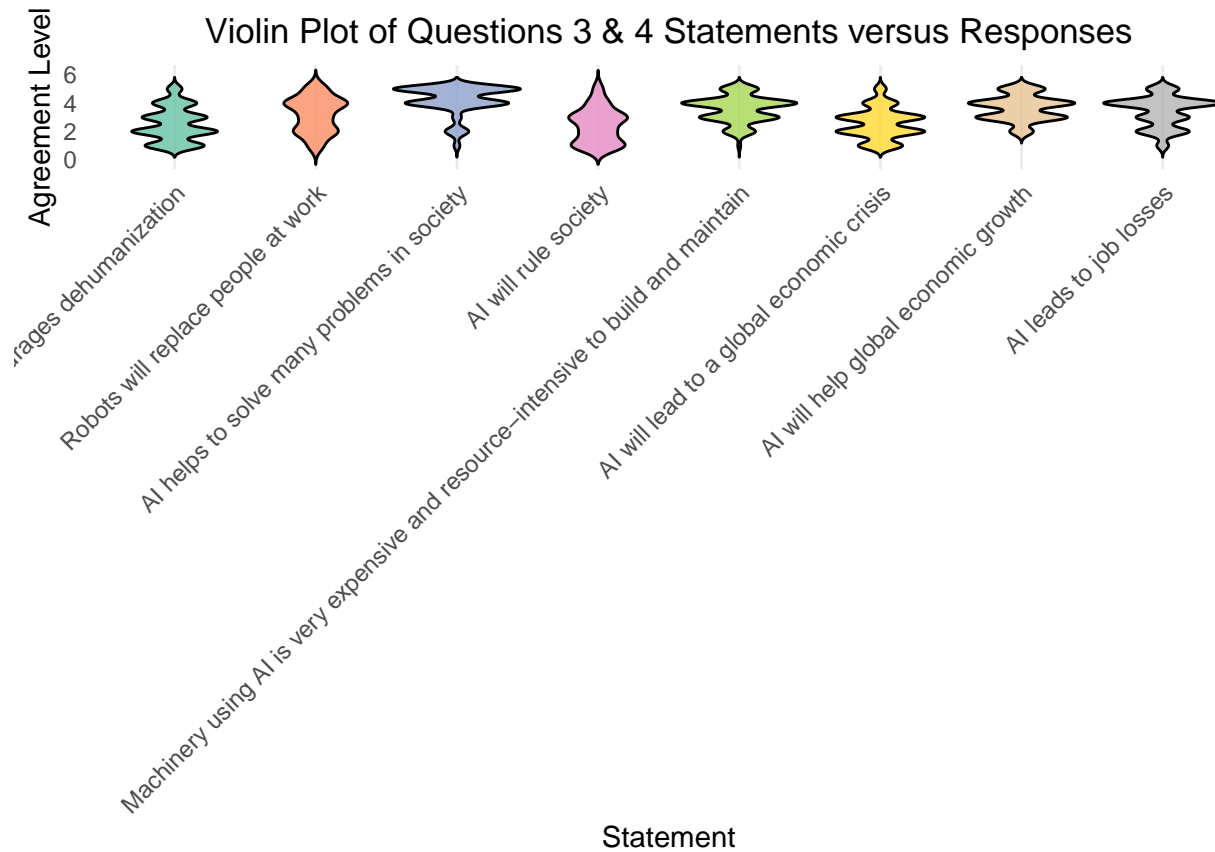
# Reshape the data for plotting
data_subset_long <- reshape2::melt(data_subset)

## No id variables; using all as measure variables
data_subset_long$variable <- factor(data_subset_long$variable, levels = columns, labels = labels)

str(data_subset_long )

## 'data.frame': 728 obs. of 2 variables:
## $ variable: Factor w/ 8 levels "AI encourages dehumanization",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ value : num 1 2 2 4 1 3 1 1 2 4 ...

my_palette <- brewer.pal(8, "Set2")
# Create a violin plot
ggplot(data_subset_long, aes(x = variable, y = value, fill = variable)) +
  geom_violin(color = "black", alpha = 0.8, scale = "count", trim = FALSE) +
  scale_fill_manual(values = my_palette) +
  labs(x = "Statement", y = "Agreement Level", fill = "") +
  ggtitle("Violin Plot of Questions 3 & 4 Statements versus Responses") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5),
        legend.position = "none",
        panel.grid.major.y = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank())
```



```
##### standard deviations in answers

# Calculate standard deviations
standard_deviations <- sapply(columns, function(column) {
  sd(data[[column]])
})

# Convert standard deviations to a data frame
sd_data <- data.frame(variable = names(standard_deviations),
                      sd = standard_deviations)

# Sort the data frame by standard deviation values
sd_data <- sd_data[order(sd_data$sd), ]

# Create a styled table
styled_table <- sd_data %>%
  kable(col.names = c("Variable", "Standard Deviation")) %>%
  kable_styling(full_width = FALSE, bootstrap_options = c("striped", "hover"))

styled_table
```

	Variable	Standard Deviation
Q4.3.Economic_growth	Q4.3.Economic_growth	0.8720319
Q4.1.AI_costly	Q4.1.AI_costly	0.9206623
Q3.3.Problem_solving	Q3.3.Problem_solving	1.0134625
Q4.2.Economic_crisis	Q4.2.Economic_crisis	1.0562911

Q4.4.Job_loss	Q4.4.Job_loss	1.0737175
Q3.1.AI_dehumanization	Q3.1.AI_dehumanization	1.1677389
Q3.4.AI_rulling_society	Q3.4.AI_rulling_society	1.2031704
Q3.2.Job_replacement	Q3.2.Job_replacement	1.2130749

```
# Extract the desired values and labels from the dataframe
feelings <- c('Curiosity', 'Fear', 'Indifference', 'Trust')
# Adjust the indices accordingly
str(data$Q5.Feelings)

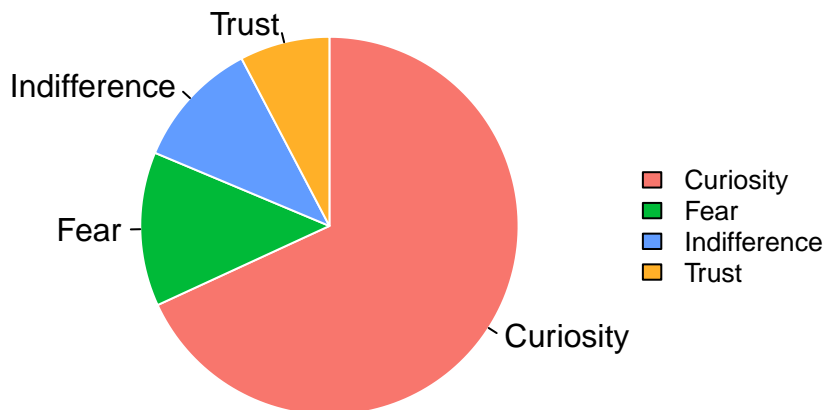
##  num [1:91] 1 1 1 1 1 2 1 1 1 1 1 ...

print(data$Q5.Feelings)

##  [1] 1 1 1 1 1 2 1 1 1 1 1 1 1 4 1 3 3 3 2 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1
## [39] 2 1 2 1 1 1 1 1 2 4 1 1 1 1 2 4 1 1 1 3 1 4 1 1 1 1 1 1 2 2 1 2 1 4 1 1 3 1
## [77] 1 1 3 3 1 2 1 3 1 2 1 1 4 2 4

counts <- table(data$Q5.Feelings)
# Define custom colors for the pie slices
custom_colors <- c('#F8766D', '#00BA38', '#619CFF', '#FFB028')

# Create the pie chart with enhanced appearance
pie(counts, labels = feelings, col = custom_colors, border = 'white',
    init.angle = 90, clockwise = TRUE)
legend("right", legend = feelings, fill = custom_colors,
    bty = "n", cex = 0.8)
```



// Weź spójnij jeszcze to Q12.Gender:

Gender of the student (1 = Male, 2 = Female) i poczytaj ten paper tam co jest, bo reszty to nie ma co zmieniać się bawić. Można jeszcze wiesz zrobić jakoś, że tyle male dało, a tyle female, ale to takie odaj te comparission tych danych do co oni mają. zrobimy korelacje, ładnie to się opisze, zrobi się prezkę i będzie gitówka myślę.