MM325M1P2

Caleb Roper

2024-04-16

Getting Started

The WVS has just completed wave 7 data that comprises 64 surveys conducted in 2017-2022. With 64 countries and societies around the world and more than 80,000 respondents, this is the latest resource made available for the research community.

Variables of Interest

My four chosen variables of interest from the World Values Survey are as follows. Detailed definitions and response categories for these variables can be found in the Codebook attached to my submission:

- **Q260 (Sex)**
- **X003R (Age)**
- **Q263 (Respondent Immigrant)**
- **Q279 (Employment Status)**

Sources

1. *World Values Survey Contents*
2. *World Values Survey Documentation WV7*
3. Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano J., M. Lagos, P. Norris, E. Ponarin & B. Puranen (eds.). 2022. *World Values Survey: Round Seven - Country-Pooled Datafile Version 5.0.* Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. doi:10.14281/18241.20

Get Data

Here I will load the data from *World Values Survey*.

```
# Load data
WVS7 <- readRDS("C:/Users/caleb/Documents/WVS_Cross-National_Wave_7_Rds_v5_0.rds")

# Convert B_COUNTRY to numeric for filtering
WVS7$B_COUNTRY <- as.numeric(as.character(WVS7$B_COUNTRY))

# Create dataset for only US respondents, variable conversions, cleaning data
us_data <- WVS7 %>%
  filter(B_COUNTRY == 840) %>%  # Filter for USA
```

```
  filter(!X003R %in% c(-1, -2, -4, -5)) %>%  # Exclude invalid age responses
  filter(!Q260 %in% c(-2, -4, -5)) %>%  # Exclude invalid sex responses
  filter(!Q263 %in% c(-1, -2, -4, -5)) %>%  # Exclude invalid immigrant status responses
  filter(!Q279 %in% c(-1, -2, -4, -5))  # Exclude invalid employment responses

# Convert necessary variables to factors with levels and labels
us_data$X003R <- factor(us_data$X003R,
                        levels = c(1, 2, 3, 4, 5, 6),
                        labels = c("16-24 years",
                                   "25-34 years",
                                   "35-44 years",
                                   "45-54 years",
                                   "55-64 years",
                                   "65 and over"))
us_data$Q279 <- factor(us_data$Q279,
                       levels = c(1, 2, 3, 4, 5, 6, 7, 8),
                       labels = c("Full time",
                                  "Part time",
                                  "Self employed",
                                  "Retired",
                                  "Housewife",
                                  "Student",
                                  "Unemployed",
                                  "Other"))
us_data$Q260 <- factor(us_data$Q260,
                       levels = c(1, 2),
                       labels = c("Male", "Female"))
us_data$Q263 <- factor(us_data$Q263,
                       levels = c(1, 2),
                       labels = c("Born in this country",
                                  "Immigrant"))
```

Data Quality Assessment

Now that the data has been loaded, we can examine the variables to ensure the sample was, in essence, broad and representative enough of the population to serve our purpose.

```
# Count of respondents from the United States
count_us_respondents <- nrow(us_data)

# Print the count of US respondents
print(paste("Count of US respondents:",
            count_us_respondents))
```

```
## [1] "Count of US respondents: 2523"
```

Due to the large amount of data, spread, and relevance of the variables being studied, we can broadly state that the dataset we are working with is reputable, broad, and representative enough of the population to serve our purpose.

Visualizations

Here I take a look at the data using some charts. The following visualizations will help highlight the distribution and relationships between variables like sex, age, immigrant status, and employment in a global context.

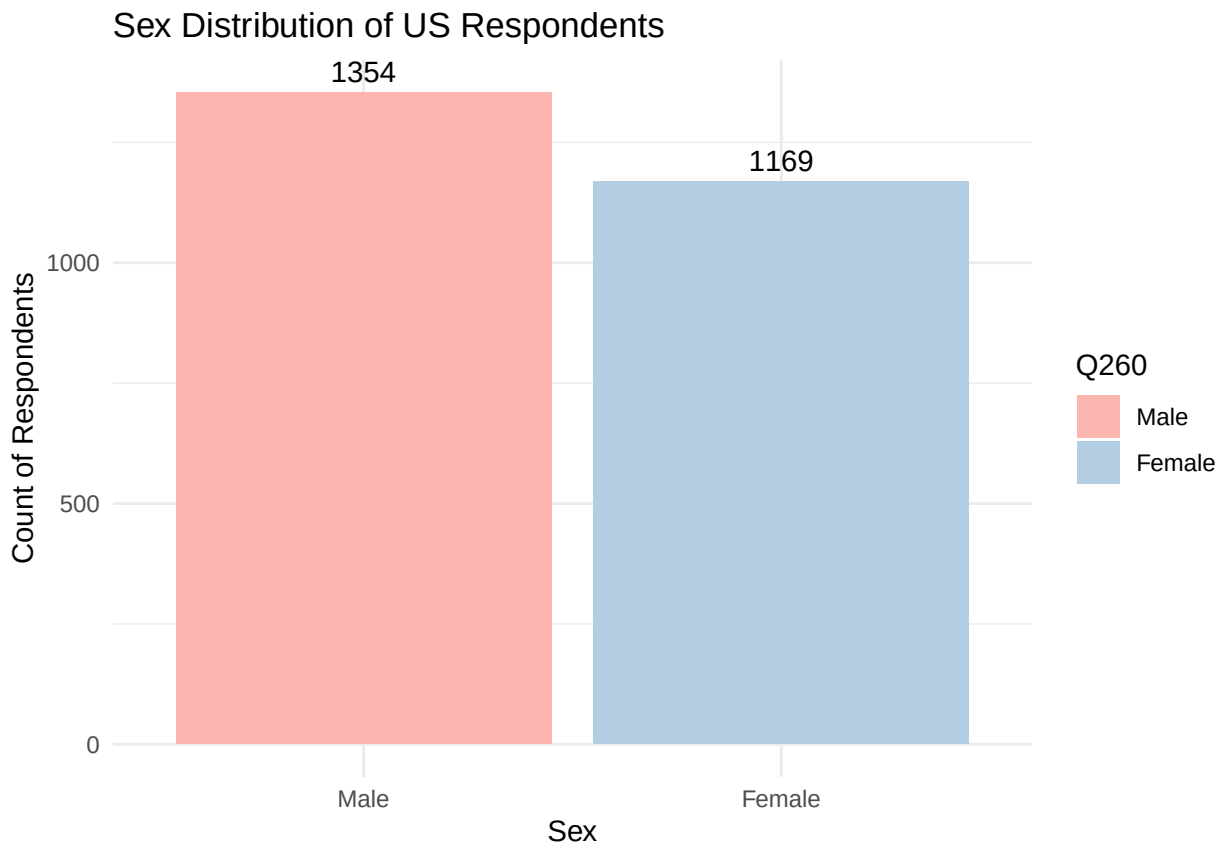
Sex

This visualization shows the count of respondents by sex, providing insight into the gender balance in the dataset.

```
# Summary of sex distribution
sex_distribution_summary <- us_data %>%
  group_by(Q260) %>%
  summarise(Count = n()) %>%
  mutate(Percentage = (Count / sum(Count)) * 100)

# Visualization of sex distribution
sex_distribution_plot <- ggplot(us_data, aes(x = Q260, fill = Q260)) +
  geom_bar(stat = "count") +
  geom_text(aes(label=after_stat(count)), stat='count', vjust=-0.5) +
  labs(title = "Sex Distribution of US Respondents",
       x = "Sex",
       y = "Count of Respondents") +
  scale_fill_brewer(palette = "Pastel1") +
  theme_minimal()

# Print results and the plot
print(sex_distribution_summary)
```

```
## # A tibble: 2 x 3
##   Q260   Count Percentage
##   <fct>  <int>      <dbl>
## 1 Male    1354       53.7
## 2 Female  1169       46.3
```

```
print(sex_distribution_plot)
```



Age

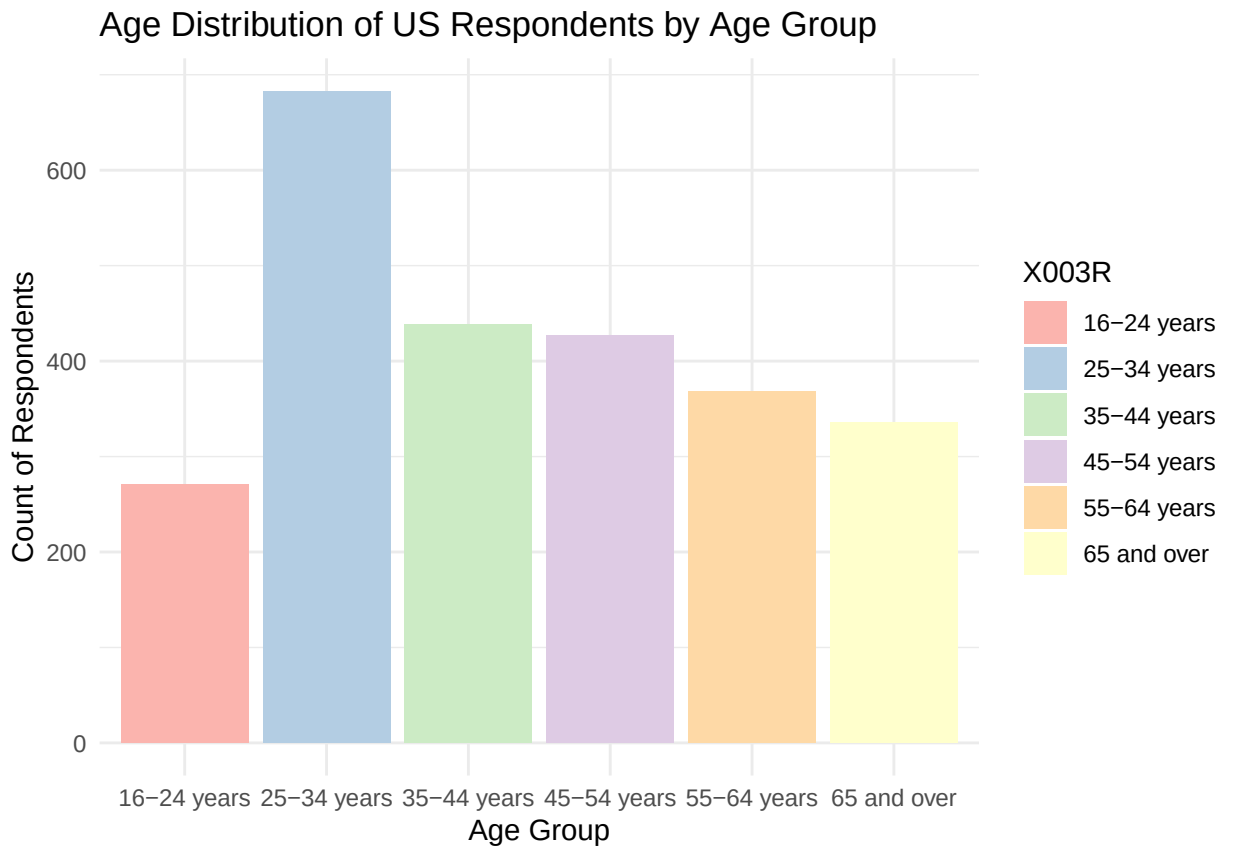
This will help us understand the age structure of the respondents, which is crucial for analyzing diversity.

```
# Summary of age distribution by category
age_distribution_us <- us_data %>%
  group_by(X003R) %>%
  summarise(Count = n()) %>%
  mutate(Percentage = (Count / sum(Count)) * 100)

# Visualize age distribution
age_plot <- ggplot(us_data, aes(x = X003R, fill = X003R)) +
  geom_bar() +
  labs(title = "Age Distribution of US Respondents by Age Group",
       x = "Age Group",
       y = "Count of Respondents") +
  scale_fill_brewer(palette = "Pastel1") +
  theme_minimal()

# Print results
print(age_distribution_us)
```

```
## # A tibble: 6 x 3
##   X003R       Count Percentage
##   <fct>       <int>      <dbl>
```

```
## 1 16-24 years   271       10.7
## 2 25-34 years   683       27.1
## 3 35-44 years   438       17.4
## 4 45-54 years   427       16.9
## 5 55-64 years   368       14.6
## 6 65 and over   336       13.3
```

```r
print(age_plot)
```



Immigrant Status

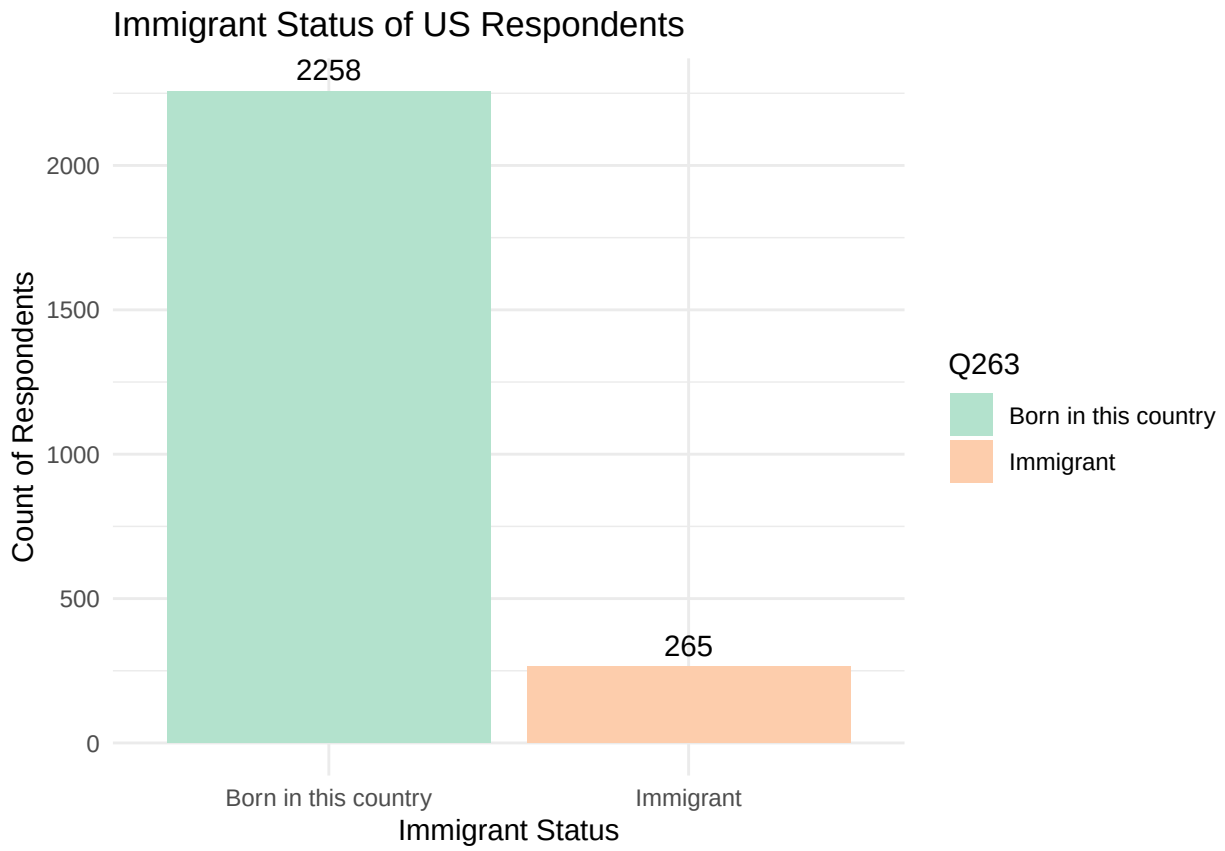
By comparing the age distribution of natives versus immigrants, we can explore differences in the age profiles between these groups.

```r
# Summary of immigrant status
immigrant_status_summary <- us_data %>%
  group_by(Q263) %>%
  summarise(Count = n()) %>%
  mutate(Percentage = (Count / sum(Count)) * 100)

# Visualization of immigrant status
immigrant_status_plot <- ggplot(us_data, aes(x = Q263, fill = Q263)) +
  geom_bar(stat = "count") +
  geom_text(aes(label = after_stat(count)), stat = 'count', vjust = -0.5) +
```

```
  labs(title = "Immigrant Status of US Respondents",
       x = "Immigrant Status",
       y = "Count of Respondents") +
  scale_fill_brewer(palette = "Pastel2") +
  theme_minimal()

# Print results and the plot
print(immigrant_status_summary)
```

```
## # A tibble: 2 x 3
##   Q263                Count Percentage
##   <fct>               <int>      <dbl>
## 1 Born in this country  2258       89.5
## 2 Immigrant              265       10.5
```

```
print(immigrant_status_plot)
```



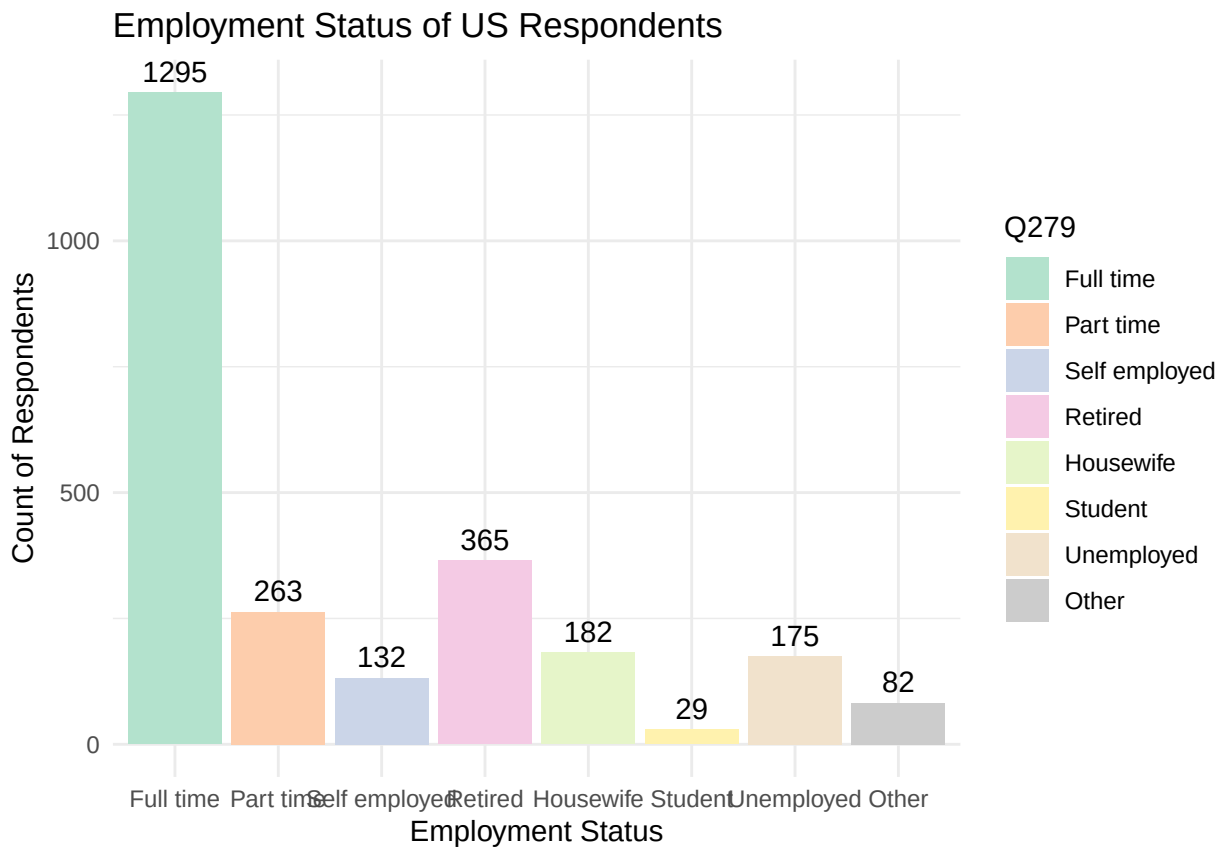
Employment Status

This will visualize how employment status varies, which can be further broken down by sex, age, or immigrant status to explore how employment is influenced by these factors.

```r
# Summary of employment status
employment_status_summary <- us_data %>%
  group_by(Q279) %>%
  summarise(Count = n()) %>%
  mutate(Percentage = (Count / sum(Count)) * 100)

# Visualization of employment status
employment_status_plot <- ggplot(us_data, aes(x = Q279, fill = Q279)) +
  geom_bar(stat = "count") +
  geom_text(aes(label = after_stat(count)), stat = 'count', vjust = -0.5) +
  labs(title = "Employment Status of US Respondents",
       x = "Employment Status",
       y = "Count of Respondents") +
  scale_fill_brewer(palette = "Pastel2") +
  theme_minimal()

# Print results and the plot
print(employment_status_summary)
```

```
## # A tibble: 8 x 3
##   Q279          Count Percentage
##   <fct>         <int>      <dbl>
## 1 Full time      1295      51.3
## 2 Part time       263      10.4
## 3 Self employed   132       5.23
## 4 Retired         365      14.5
## 5 Housewife       182       7.21
## 6 Student          29       1.15
## 7 Unemployed      175       6.94
## 8 Other            82       3.25
```

```r
print(employment_status_plot)
```



Memo to the Head of Cultural Affairs

Overview

The following memo is divided into 3 sections and includes additional insights based on assingment requirements and the data analysis thus far.

What you are studying and why.

In this analysis we are studying the World Values Survey Contents (WVS7) to determine its potential insights for members of the board. Since my team has identified the World Values Survey as a potentially useful data source to leverage a broad range of cultural insights for the project, our analysis uses it as a primary source. We are studying this to better understand how we can effectively serve and provide services to our target markets.

List the four variables, including the column name for each, as is shown in the data, and a brief description indicating what each variable represents.

1. **Q260 (Sex)**
   Description: This is the respondent's sex.
   Permitted Values:

- 1.- Male

- 2.- Female
- -2-.- No answer
- -4-.- Not asked
- -5-.- Missing; Not available; AU: Other

3. **X003R (Age): Age Recorded (6 intervals)**
   Description: This is the respondent's age as recorded in six different intervals.
   Permitted Values:

- 1.- 16-24 years
- 2.- 25-34 years
- 3.- 35-44 years
- 4.- 45-54 years
- 5.- 55-64 years
- 6.- 65 and over
- -1-.- Don't know
- -2-.- No answer
- -4-.- Not asked
- -5-.- Missing; Not available; AU: Other

3. **Q263 (Respondent Immigrant)**
   Description: This is a respondents answer to the question, were you born in this country or are you an immigrant?
   Permitted Values:

- 1.- I am born in this country
- 2.- I am an immigrant to this country (born outside this country)
- -1-.- Don´t know
- -2-.- No answer
- -4-.- Not asked
- -5-.- Missing; Not available

4. **Q279 (Employment Status)**
   Description: This is a respondents answer to the question, are you employed now or not? If yes: About how many hours a week?
   Permitted Values:

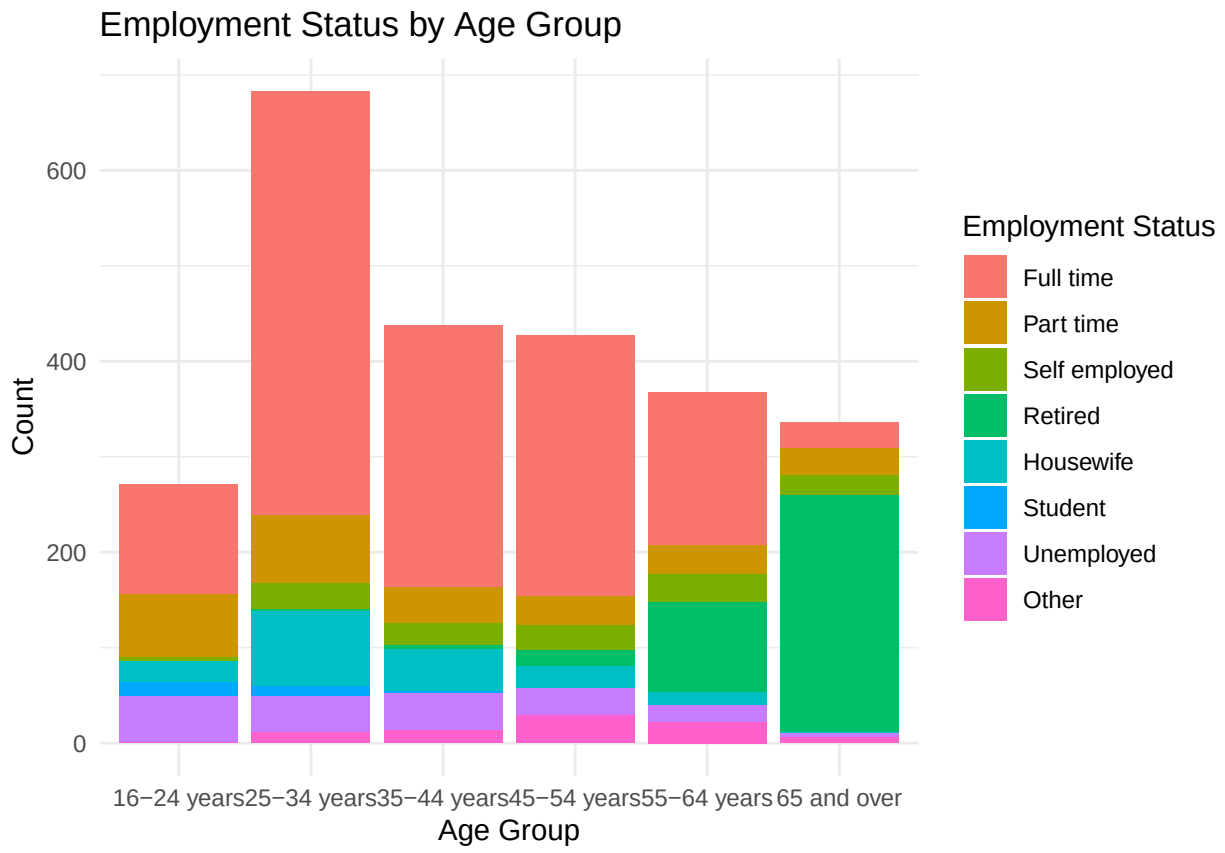
- 1.- Full time (30 hours a week or more)
- 2.- Part time (less than 30 hours a week)
- 3.- Self employed
- 4.- Retired/pensioned
- 5.- Housewife not otherwise employed
- 6.- Student
- 7.- Unemployed
- 8.- Other
- -1-.- Don´t know
- -2-.- No answer
- -4-.- Not asked
- -5-.- Missing; Not available

How can you use those variables to assess the value of multiculturalism and diversity in a global environment? What types of visualizations can be used to assist with that assessment?

Based on the previous 4 charts we can consider combining these variables to gain deeper insights into the data. These combinations highlight interactions between demographic characteristics and socio-cultural trends. Below are 3 additional charts that blend these variables and provide deeper insights for the head of cultural affairs to consider.

Age and Employment Status

```
# Plot
ggplot(us_data, aes(x = X003R, fill = Q279)) +
  geom_bar(position = "stack") +
  labs(title = "Employment Status by Age Group",
       x = "Age Group",
       y = "Count",
       fill = "Employment Status") +
  theme_minimal()
```



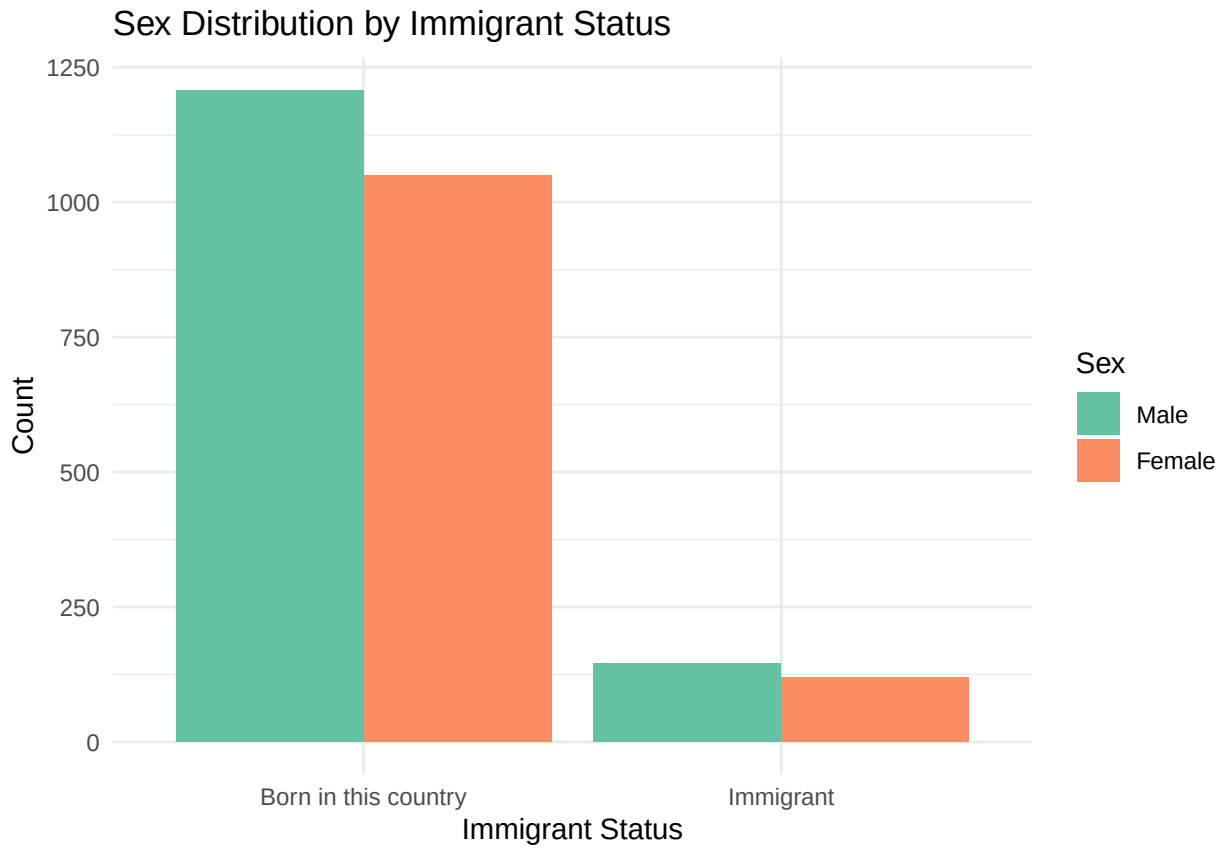

```
# Create a summary table
age_employment_summary <- us_data %>%
  group_by(X003R, Q279) %>%
  summarise(Count = n(), .groups = 'drop') %>%
  arrange(X003R, Q279)

# Print the table using knitr::kable
knitr::kable(age_employment_summary, caption = "Counts of Employment Status by Age Group")
```

Table 1: Counts of Employment Status by Age Group

X003R	Q279	Count
16-24 years	Full time	115
16-24 years	Part time	66
16-24 years	Self employed	5
16-24 years	Housewife	21
16-24 years	Student	15
16-24 years	Unemployed	48
16-24 years	Other	1
25-34 years	Full time	444
25-34 years	Part time	71
25-34 years	Self employed	28
25-34 years	Retired	2
25-34 years	Housewife	78
25-34 years	Student	11
25-34 years	Unemployed	38
25-34 years	Other	11
35-44 years	Full time	275
35-44 years	Part time	37
35-44 years	Self employed	23
35-44 years	Retired	4
35-44 years	Housewife	44
35-44 years	Student	3
35-44 years	Unemployed	39
35-44 years	Other	13
45-54 years	Full time	273
45-54 years	Part time	30
45-54 years	Self employed	27
45-54 years	Retired	16
45-54 years	Housewife	23
45-54 years	Unemployed	29
45-54 years	Other	29
55-64 years	Full time	161
55-64 years	Part time	30
55-64 years	Self employed	29
55-64 years	Retired	95
55-64 years	Housewife	14
55-64 years	Unemployed	17
55-64 years	Other	22
65 and over	Full time	27
65 and over	Part time	29
65 and over	Self employed	20
65 and over	Retired	248
65 and over	Housewife	2
65 and over	Unemployed	4
65 and over	Other	6

Sex and Immigrant Status

```r
# Plot
ggplot(us_data, aes(x = Q263, fill = Q260)) +
  geom_bar(position = "dodge") +
  labs(title = "Sex Distribution by Immigrant Status",
       x = "Immigrant Status",
       y = "Count",
       fill = "Sex") +
  scale_fill_brewer(palette = "Set2") +
  theme_minimal()
```




```r
# Create a summary table for Sex and Immigrant Status
sex_immigrant_summary <- us_data %>%
  group_by(Q263, Q260) %>%
  summarise(Count = n(), .groups = 'drop') %>%
  arrange(Q263, Q260)

# Print the table using knitr::kable
knitr::kable(sex_immigrant_summary, caption = "Counts of Respondents by Sex and Immigrant Status")
```

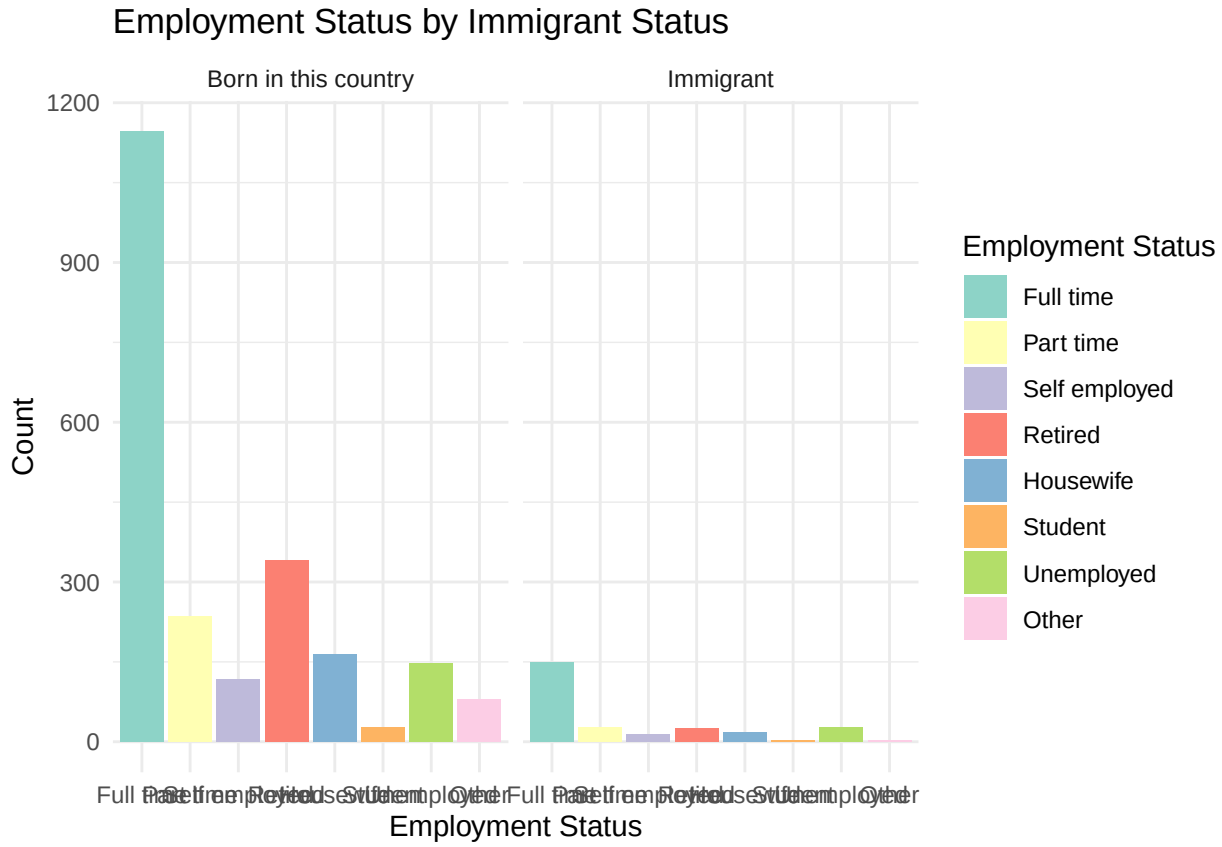
Table 2: Counts of Respondents by Sex and Immigrant Status

Q263	Q260	Count
Born in this country	Male	1208
Born in this country	Female	1050

Q263	Q260	Count
Immigrant	Male	146
Immigrant	Female	119

Employment Status and Immigrant Status

```
# Plot
ggplot(us_data, aes(x = Q279, fill = Q279)) +
  geom_bar(stat = "count") +
  facet_wrap(~ Q263) +
  labs(title = "Employment Status by Immigrant Status",
       x = "Employment Status",
       y = "Count",
       fill = "Employment Status") +
  scale_fill_brewer(palette = "Set3") +
  theme_minimal()
```




```
# Create a summary table for Employment Status and Immigrant Status
employment_immigrant_summary <- us_data %>%
  group_by(Q263, Q279) %>%
  summarise(Count = n(), .groups = 'drop') %>%
  arrange(Q263, Q279)

# Print the table using knitr::kable
```

```
knitr::kable(employment_immigrant_summary,
             caption = "Counts of Employment Status by Immigrant Status")
```

Table 3: Counts of Employment Status by Immigrant Status

Q263	Q279	Count
Born in this country	Full time	1146
Born in this country	Part time	236
Born in this country	Self employed	118
Born in this country	Retired	340
Born in this country	Housewife	164
Born in this country	Student	27
Born in this country	Unemployed	148
Born in this country	Other	79
Immigrant	Full time	149
Immigrant	Part time	27
Immigrant	Self employed	14
Immigrant	Retired	25
Immigrant	Housewife	18
Immigrant	Student	2
Immigrant	Unemployed	27
Immigrant	Other	3

Insights and Recommendations for the Head of Cultural Affairs:

1. Diverse Employment Needs: Full-time employment is most common, but significant numbers are in part-time, self-employed, and unemployed categories. I would advise to create culturally relevant programs that accommodate varied work schedules and support entrepreneurial ventures.

2. Gender and Immigrant Demographics: There are more males in both native and immigrant groups, with a significantly smaller immigrant population. I would advise enhancing gender equality initiatives and cultural integration programs tailored to both demographics, focusing on language support and community engagement.

3. Employment Among Natives and Immigrants: Natives show higher employment figures across all categories compared to immigrants, with noticeable challenges in immigrant unemployment and underrepresentation in education and retirement. I would advise development of specific economic integration programs for immigrants, bolsteing educational access, and ensuring retirement support is adequate.

Overall, the head of cultural affairs should prioritize inclusive programming that addresses the diverse cultural and economic backgrounds of the community, ensuring equal opportunities for engagement and support for all groups.