````
---
title: "ECON 121 FA23 Problem Set 1"
author: "Robert Tso"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE) # sets the code chunk format
rm(list = ls()) # clears the environment
```

## Question 1

Verbal: list group members.
Robert Tso

\newpage

## Question 2

Code: Load packages and dataset, summarize the data.

Verbal: Interpret the summary statistics.

```{r message=FALSE, warning=FALSE}
# The PDF will show the code you write here but not the output.
# Load packages and dataset here.
library(readr)

install.packages("tidyverse")
library(tidyverse)
library(dplyr)
library(readr)
ssa_names <- read_csv("D:/Documents/Class/Econ 121/econ121/data/ssa_names.csv")
View(ssa_names)

```

```{r echo=TRUE}
# The PDF will show the code AND output here.
# Summarize the data here.
# List of newborn boys and girls starting from 1940 to 2022.

summarise(ssa_names)



```

\newpage

## Question 3

Code: Find the all time most common boy and girl names.

Verbal: Interpret your results.

```{r echo=TRUE}

#aggregate by name
total_names <- ssa_names%>%
  group_by(sex,name) %>%
  summarize(total_frequency = sum(frequency))%>%
  ungroup()
````
````

```
total_names

#girls
total_girl_names <- total_names%>%
  filter(sex=="F") #filter by sex for female


total_girl_names%>%
  filter(total_frequency==max(total_frequency)) #filter total girl names by the most
frequent name


most_common_girl <- max(total_girl_names$total_frequency) #store the frequency for girls



#boys
total_boy_names <- total_names%>%
  filter(sex=="M") #filter by sex for male


total_boy_names%>%
  filter(total_frequency==max(total_frequency)) #filter total boy names by the most
frequent name


most_common_boy <- max(total_boy_names$total_frequency) #store the frequency for boy

#add the total frequencies
print(most_common_girl+most_common_boy)
#Total of 6,261,170 babies of the 2 most popular names.

#Mary was the most popular name for girls at 1,997,025 and Michael for boys at 4,264,145 A
total of 6,261,170 babies since 1940 were born under these 2 popular names.
```

\newpage

Question 4

Code: Plot time trends in the number of unique names by sex

Verbal: Interpret your results.

```{r echo=TRUE}
# All question 4 code here

unique_girls_by_year <- ssa_names %>%
  filter(sex=="F")%>%
  group_by(year) %>%
  summarise(unique_name_count = n_distinct(name)) #count the amount of distinct names each
year for girls

unique_girls_by_year

unique_boys_by_year <- ssa_names %>%
  filter(sex=="M")%>%
  group_by(year) %>%
  summarise(unique_name_count = n_distinct(name))#count the amount of distinct names each
year for boys

#unique_boys_by_year
```

```
unique_names_plot <- ggplot() +
  geom_line(data = unique_boys_by_year, aes(x = year, y = unique_name_count, color =
"Boys")) +
  geom_line(data = unique_girls_by_year, aes(x = year, y = unique_name_count, color =
"Girls")) +

  labs(x = "Year", y = "Number of Unique Names", title = "Unique Names by Year (Boys vs.
Girls)") +
  scale_color_manual(values = c("Boys" = "blue", "Girls" = "salmon")) +
  theme_minimal()

unique_names_plot

#Throughout the years, girls have had more unique names than boys, and they trend in
similar ways. The number of unique names for both boys has remained relatively flat
compared to girls until 1960-1970s. Both sexes peaked in uniqueness in 2008, where boys
returned to a flat rate, while girls started declining.

```

\newpage

Question 5

Code: Plot time trends in the number of unique names relative to the number of
babies by sex.

Verbal: Interpret your results.

```{r echo=TRUE}
# All question 5 code here

total_girl_babies <- ssa_names %>%
  filter(sex=="F")%>%
  group_by(year) %>%
  summarise(ratio = n_distinct(name)/sum(frequency))
total_girl_babies

total_boy_babies <- ssa_names %>%
  filter(sex=="M")%>%
  group_by(year) %>%
  summarise(ratio = n_distinct(name)/sum(frequency))
total_boy_babies


ratio_names_plot <- ggplot() +
  geom_line(data = total_boy_babies, aes(x = year, y = ratio, color = "Boys")) +
  geom_line(data = total_girl_babies, aes(x = year, y = ratio, color = "Girls")) +

  labs(x = "Year", y = "Ratio", title = "Ratio of Unique Names by Year (Boys vs. Girls)")
+
  scale_color_manual(values = c("Boys" = "blue", "Girls" = "salmon")) +
  theme_minimal()

ratio_names_plot

#The ratio of unique names increases over time, with slight declines during certain time
periods, such as the mid-1970s to mid-1980s, and after 2010, for girls a steady decline
while for boys only a short decline. The patterns display a decreasing of conformity among
babies.

```

\newpage

Question 6

Verbal: Which names and why?

I chose Apple because it was just the first word I could think of that starts with "A". Sam because this is a gender neutral name, short for Samuel or Samantha.

\newpage

Question 7

Code: Find all-time frequency of your names and compare with all-time most popular names.

Verbal: Interpret your results.

```{r echo=TRUE}
# All question 7 code here

#apple
Apple_names <- ssa_names%>%
  filter(name=="Apple")
Apple_names_girls <- Apple_names%>%
  filter(sex=="F")
Apple_names_boys <- Apple_names%>%
  filter(sex=="M")

#Apple_names_girls
#Apple_names_boys


Sam_names <- ssa_names %>%
  filter(name=="Sam")

Apple_names_plot_log <- ggplot() +
  geom_line(data = unique_boys_by_year, aes(x = year, y = unique_name_count, color =
"Boys")) +
  geom_line(data = unique_girls_by_year, aes(x = year, y = unique_name_count, color =
"Girls")) +
  geom_line(data = Apple_names_boys, aes(x = year, y = frequency, color = "Apple Boys")) +
  geom_line(data = Apple_names_girls, aes(x = year, y = frequency, color = "Apple
Girls"))+

  scale_y_continuous(trans='log10')+           #Scaled the graph into Log
  labs(x = "Year", y = "Number of Unique Names", title = "Unique Names by Year (Boys vs.
Girls)") +
  scale_color_manual(values = c("Boys" = "blue", "Girls" = "salmon","Apple
Boys"="purple","Apple Girls"="red")) +
  theme_minimal()

#sam
Sam_names <- ssa_names%>%
  filter(name=="Sam")
Sam_names_girls <- Sam_names%>%
  filter(sex=="F")
Sam_names_boys <- Sam_names%>%
  filter(sex=="M")

#Sam_names_girls
#Sam_names_boys



Sam_names_plot_log <- ggplot() +
```

```
  geom_line(data = unique_boys_by_year, aes(x = year, y = unique_name_count, color =
"Boys")) +
  geom_line(data = unique_girls_by_year, aes(x = year, y = unique_name_count, color =
"Girls")) +
  geom_line(data = Sam_names_boys, aes(x = year, y = frequency, color = "Sam Boys")) +
  geom_line(data = Sam_names_girls, aes(x = year, y = frequency, color = "Sam Girls"))+

  scale_y_continuous(trans='log10')+           #Scaled the graph into Log
  labs(x = "Year", y = "Number of Unique Names", title = "Unique Names by Year (Boys vs.
Girls)") +
  scale_color_manual(values = c("Boys" = "blue", "Girls" = "salmon","Sam
Boys"="green","Sam Girls"="orange")) +
  theme_minimal()

Sam_names_plot <- ggplot() +
  geom_line(data = unique_boys_by_year, aes(x = year, y = unique_name_count, color =
"Boys")) +
  geom_line(data = unique_girls_by_year, aes(x = year, y = unique_name_count, color =
"Girls")) +
  geom_line(data = Sam_names_boys, aes(x = year, y = frequency, color = "Sam Boys")) +
  geom_line(data = Sam_names_girls, aes(x = year, y = frequency, color = "Sam Girls"))+

  labs(x = "Year", y = "Number of Unique Names", title = "Unique Names by Year (Boys vs.
Girls)") +
  scale_color_manual(values = c("Boys" = "blue", "Girls" = "salmon","Sam
Boys"="green","Sam Girls"="orange")) +
  theme_minimal()

Apple_names_plot_log
#I had not checked beforehand how many boys were named Apple, so I was surprised to see no
line representing boys named Apple

Sam_names_plot_log
#I was surprised to see how erratic, but constant Sam's name was for girls, meanwhile for
boys, Sam was a slow decline.
Sam_names_plot
#I checked again without the Log, and see that Sam is near non-existent among girls, while
Sam was still relatively popular before the 1960s for boys.

```

\newpage

Question 8

Code: Graph the annual frequencies of your chosen name(s) over time.

Verbal: Interpret your results.

```{r echo=TRUE}
# All question 8 code here

picked_names_plot <- ggplot() +
  geom_line(data = Apple_names_boys, aes(x = year, y = frequency, color = "Apple Boys")) +
  geom_line(data = Apple_names_girls, aes(x = year, y = frequency, color = "Apple
Girls"))+
  geom_line(data = Sam_names_boys, aes(x = year, y = frequency, color = "Sam Boys")) +
  geom_line(data = Sam_names_girls, aes(x = year, y = frequency, color = "Sam Girls"))+

  scale_y_continuous(trans='log10')+           #Scaled the graph into Log
  labs(x = "Year", y = "Number of Unique Names", title = "Unique Names by Year (Boys vs.
Girls)") +
  scale_color_manual(values = c("Apple Boys"="purple","Apple Girls"="red","Sam
Boys"="green","Sam Girls"="orange")) +
  theme_minimal()
```

```
picked_names_plot
#Sam and Apple are similar in popularity among girls, as they are very unconventional names, and Apple only started appearing after the 1970s, predictably more popular among hippie/non-conformist parents. Sam as a name for boys sees a steady decline after 1960, which is the same time as the near exponential increase of non-conformity among all baby names mentioned in question 5.

````