R Notebook

Code

Hide

```
#Loading relevant packages
library(tidyverse)
library(dplyr)
library(ggplot2)
library(scales)

#Loading the data into RStudio
citizenship <- read.csv("C:\\Users\\sophi\\OneDrive\\Documents\\R Projects\\granted_citizenship_1949_to_2019.csv")

#First we want to get rid of any NA values, and replace with 0 in order to prevent plotting issues.
citizenship[is.na(citizenship)] <- 0
head(citizenship)
```

	Country.of.Birth <chr>	Total <int>	X. <chr>	X1949 <dbl>	X1950 <dbl>	X1951 <dbl>	X1952 <dbl>	X1953 <dbl>	X1954 <dbl>
1	Aden	25	0.00%	0	0	0	0	0	0
2	Afghanistan	3499	0.38%	0	0	0	0	0	0
3	Albania	144	0.02%	0	0	0	0	0	0
4	Algeria	241	0.03%	0	0	0	0	0	0
5	American Samoa	862	0.09%	0	0	0	0	0	0
6	Angola	26	0.00%	0	0	0	0	0	0

6 rows | 1-10 of 74 columns

Hide

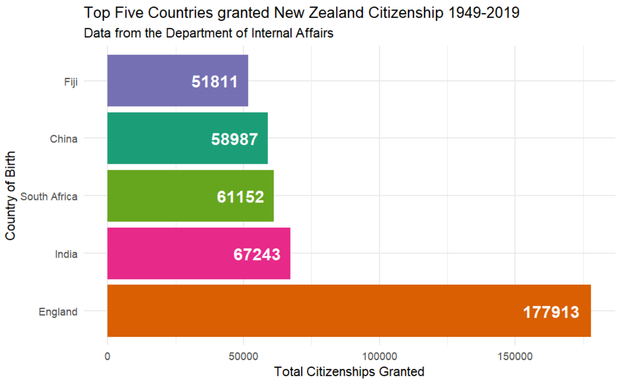
```
#Let's also remove the row with TOTAL amount of citizenships.
citizenship <- filter(citizenship, Country.of.Birth != "TOTAL", .preserve = FALSE)


#Let's find the top 5 countries that people who earned NZ citizenship were born in.
top_countries <- head(arrange(citizenship, desc(Total)), 5)
top_countries
```

	Country.of.Birth <chr>	Total <int>	X. <chr>	X1949 <dbl>	X1950 <dbl>	X1951 <dbl>	X1952 <dbl>	X1953 <dbl>	X1954 <dbl>
1	England	177913	19.55%	9429	21	81	116	96	131
2	India	67243	7.39%	320	21	28	17	25	50
3	South Africa	61152	6.72%	140	0	0	6	3	1
4	China	58987	6.48%	51	1	3	4	10	12
5	Fiji	51811	5.69%	221	0	4	3	4	7

5 rows | 1-10 of 74 columns

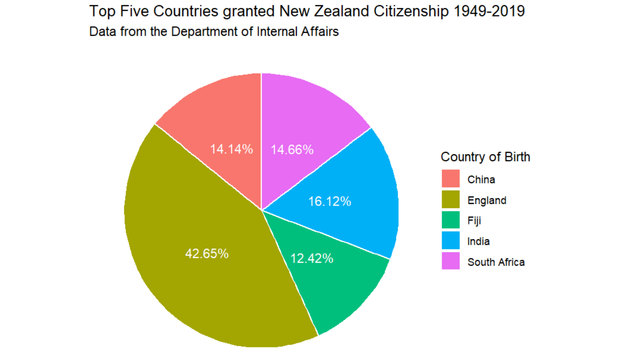
Hide

```
# Create a horizontal bar plot of top 5 countries
ggplot(data = top_countries, aes(x = fct_reorder(Country.of.Birth, -Total), y = Total, fill = Country.of.Birth))
+
  geom_bar(stat='identity') +
  theme_minimal() +
  labs(x = "Country of Birth", y = "Total Citizenships Granted",
       title ="Top Five Countries granted New Zealand Citizenship 1949-2019",
       subtitle = "Data from the Department of Internal Affairs") +
  theme(legend.position = "none") +
  geom_text(aes(label = Total), hjust = 1.2, size = 5, fontface = "bold", color = "white") +
  scale_fill_brewer(palette = "Dark2") +
  coord_flip()
```



Hide

```
# Create a piechart of top 5 countries
total_top <- sum(top_countries$Total)
top_percentage <- percent(top_countries$Total/total_top)
ggplot(data = top_countries, aes(x = "", y = Total, fill=Country.of.Birth)) +
  geom_bar(stat='identity', width=1, color="white") +
  coord_polar("y", start=0) +
  theme_void() +
  labs(title ="Top Five Countries granted New Zealand Citizenship 1949-2019",
  subtitle = "Data from the Department of Internal Affairs",
  color = "Country of Birth") +
  geom_text(aes(label = top_percentage),
  position = position_stack(vjust = 0.5), color = "white") +
  guides(fill = guide_legend(title = "Country of Birth"))
```



Hide

```
# Create a line plot of each country over time
ggplot(data = top_countries_long, aes(x = year, y = value, group=Country.of.Birth)) +
  geom_line(aes(color=Country.of.Birth), stat='identity', size=1) +
  scale_color_discrete() +
  theme_classic() +
  labs(x = "Year", y = "Number of Citizenships",
       title =  "Trends in Top Five Countries of Birth for New Zealand Citizenships, 1949-2019",
       subtitle = "Data from the Department of Internal Affairs",
       color = "Country of Birth") +
  theme(plot.title = element_text(size = 12, face = "bold", hjust = 0.5),
        plot.subtitle = element_text(size = 12, hjust = 0.5),
        axis.title = element_text(size = 14),
        axis.text = element_text(size = 12),
        legend.title = element_text(size = 14),
        legend.text = element_text(size = 12),
        legend.position = "bottom",
        legend.box.background = element_rect(color = "black", size = 1))
```