

# NZ\_Migration\_Analysis

ABBAS Hamza

2023-11-24

In this project, we will analyze the New Zeland migration dataset in order to draw conclusions and discover the patters of the insights if we got a huge deffirence between the Departures and the Arrivals

## Installing and loading packages

```
install.packages("tidyverse")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)

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    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## 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
```

## Importing the dataset from a excel file

```
x <- read_csv("/cloud/project/migration_nz.csv")

## Rows: 86526 Columns: 5
## -- Column specification -----
## Delimiter: ","
## chr (3): Measure, Country, Citizenship
## dbl (2): Year, Value
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## Processing Data

This step is key in order to eliminate the duplicates, blank\_rows and extra spaces, this step help me as an analyst to conserve the integrity of the data

```
x1 <- x %>%
  drop_na() %>%
```

```
mutate_all(trimws) %>%
distinct()
head(x1)
```

```
## # A tibble: 6 x 5
##   Measure Country   Citizenship      Year Value
##   <chr>    <chr>    <chr>          <chr> <chr>
## 1 Arrivals Oceania   New Zealand Citizen  1979 11817
## 2 Arrivals Oceania   Australian Citizen   1979  4436
## 3 Arrivals Oceania   Total All Citizenships 1979 19965
## 4 Arrivals Antarctica New Zealand Citizen   1979   10
## 5 Arrivals Antarctica Australian Citizen   1979    0
## 6 Arrivals Antarctica Total All Citizenships 1979   13
```

## More processing

```
x1 <- x1 %>%
  mutate(Value = as.integer(Value))
head(x1)
```

```
## # A tibble: 6 x 5
##   Measure Country   Citizenship      Year Value
##   <chr>    <chr>    <chr>          <chr> <int>
## 1 Arrivals Oceania   New Zealand Citizen  1979 11817
## 2 Arrivals Oceania   Australian Citizen   1979  4436
## 3 Arrivals Oceania   Total All Citizenships 1979 19965
## 4 Arrivals Antarctica New Zealand Citizen   1979    10
## 5 Arrivals Antarctica Australian Citizen   1979     0
## 6 Arrivals Antarctica Total All Citizenships 1979    13
```

```
str(x1)
```

```
## tibble [86,454 x 5] (S3: tbl_df/tbl/data.frame)
##  $ Measure      : chr [1:86454] "Arrivals" "Arrivals" "Arrivals" "Arrivals" ...
##  $ Country      : chr [1:86454] "Oceania" "Oceania" "Oceania" "Antarctica" ...
##  $ Citizenship: chr [1:86454] "New Zealand Citizen" "Australian Citizen" "Total All Citizenships" "N
##  $ Year         : chr [1:86454] "1979" "1979" "1979" "1979" ...
##  $ Value        : int [1:86454] 11817 4436 19965 10 0 13 17 4 30 8224 ...
```

## Analysis

Those calculations are to see the difference between the traffic

```
x2 <- x1 %>%
  filter(Measure == "Arrivals") %>%
  summarise(Average_Value = mean(Value))
print(x2)
```

```
## # A tibble: 1 x 1
##   Average_Value
##           <dbl>
## 1           382.
```

```
x3 <- x1 %>%
  filter(Measure == "Departures") %>%
```

```
summarise(Average_Value = mean(Value))
print(x3)
```

```
## # A tibble: 1 x 1
##   Average_Value
##           <dbl>
## 1           429.
```

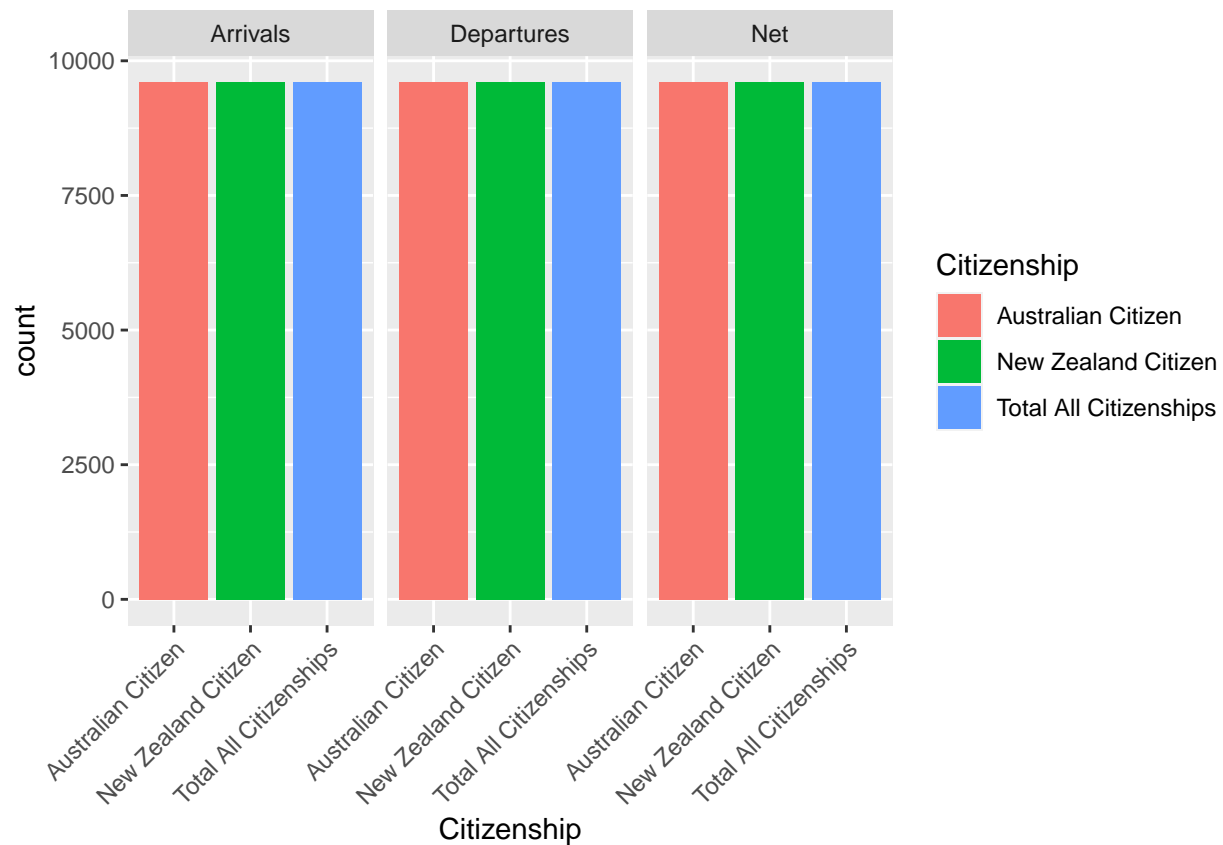
```
head(x1)
```

```
## # A tibble: 6 x 5
##   Measure Country      Citizenship      Year Value
##   <chr>    <chr>      <chr>      <chr> <int>
## 1 Arrivals Oceania    New Zealand Citizen  1979  11817
## 2 Arrivals Oceania    Australian Citizen   1979   4436
## 3 Arrivals Oceania    Total All Citizenships 1979  19965
## 4 Arrivals Antarctica New Zealand Citizen   1979     10
## 5 Arrivals Antarctica Australian Citizen   1979      0
## 6 Arrivals Antarctica Total All Citizenships 1979     13
```

## Visualization

The goal of any visualization is to effectively communicate information, patterns, or insights in a clear and accessible manner. Here are some overarching goals that apply to various types of visualizations.

```
library(ggplot2)
v <- ggplot(data = x1) +
  geom_bar(mapping = aes(x = Citizenship, fill = Citizenship)) +
  facet_grid(~Measure) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
print(v)
```



## Conclusion

As the graph tell, The departures and arrivals are very similar, which is beneficial for the demographic stability of the country. Also: Cultural Diversity: A balance between departures and arrivals can lead to a diverse and multicultural society. People arriving bring different cultural backgrounds, perspectives, and skills, enriching the overall cultural fabric of the country.

Government Planning: Government agencies can plan more effectively for public services, infrastructure, and social programs when demographic changes are gradual and predictable. This includes education, healthcare, and housing planning.

Tourism Opportunities: A balance between departures and arrivals may create a stable environment for tourism. Tourist destinations can plan for consistent visitor numbers, leading to sustainable tourism practices.