# Module 5 - Assignment 1

## Pierce, Michaela

### Data Wrangling

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readxl)  
  
UN\_migrant <- read\_excel("UN\_migrant.xlsx",   
sheet = "Table 6", col\_types = c("numeric",   
"text", "text", "numeric", "text",   
"numeric", "numeric", "numeric",   
"numeric", "numeric", "numeric","text",   
"text","text","text","text","text","text",   
"text","text","text","text"), skip = 15)

## Warning: Expecting numeric in F40 / R40C6: got '..'

## Warning: Expecting numeric in G40 / R40C7: got '..'

## Warning: Expecting numeric in H40 / R40C8: got '..'

## Warning: Expecting numeric in I40 / R40C9: got '..'

## Warning: Expecting numeric in F179 / R179C6: got '..'

## Warning: Expecting numeric in G179 / R179C7: got '..'

## Warning: Expecting numeric in H179 / R179C8: got '..'

## Warning: Expecting numeric in I179 / R179C9: got '..'

## Warning: Expecting numeric in F220 / R220C6: got '..'

## Warning: Expecting numeric in G220 / R220C7: got '..'

## Warning: Expecting numeric in H220 / R220C8: got '..'

## New names:  
## • `` -> `...1`  
## • `` -> `...2`  
## • `` -> `...3`  
## • `` -> `...4`  
## • `` -> `...5`  
## • `1990` -> `1990...6`  
## • `1995` -> `1995...7`  
## • `2000` -> `2000...8`  
## • `2005` -> `2005...9`  
## • `2010` -> `2010...10`  
## • `2015` -> `2015...11`  
## • `1990` -> `1990...12`  
## • `1995` -> `1995...13`  
## • `2000` -> `2000...14`  
## • `2005` -> `2005...15`  
## • `2010` -> `2010...16`  
## • `2015` -> `2015...17`

### Part 2 - Cleaning Data with dplyr

# Rename columns  
UN\_migrant <- UN\_migrant %>%  
 rename(Country = ...2,  
 Country\_Code = ...4,  
 Type = ...5,  
 `1990` = `1990...6`,  
 `1995` = `1995...7`,  
 `2000` = `2000...8`,  
 `2005` = `2005...9`,  
 `2010` = `2010...10`,  
 `2015` = `2015...11`)  
  
# Create new dataset  
Migration <- UN\_migrant %>%  
 select(Country, Country\_Code, Type, `1990`, `1995`, `2000`, `2005`, `2010`, `2015`)

### Part 3 - Creating tidy data using tidyr

# Create new dataset to change year to numeric  
Migration2 <- pivot\_longer(Migration, cols = -c(Country, Country\_Code, Type), names\_to = "Year", values\_to = "Cases")  
head(Migration2)

## # A tibble: 6 × 5  
## Country Country\_Code Type Year Cases  
## <chr> <dbl> <chr> <chr> <dbl>  
## 1 WORLD 900 <NA> 1990 18836571  
## 2 WORLD 900 <NA> 1995 17853840  
## 3 WORLD 900 <NA> 2000 15827803  
## 4 WORLD 900 <NA> 2005 13276733  
## 5 WORLD 900 <NA> 2010 15370755  
## 6 WORLD 900 <NA> 2015 19577474

# change year to numeric  
Migration2$Year <- as.numeric(Migration2$Year)  
head(Migration2)

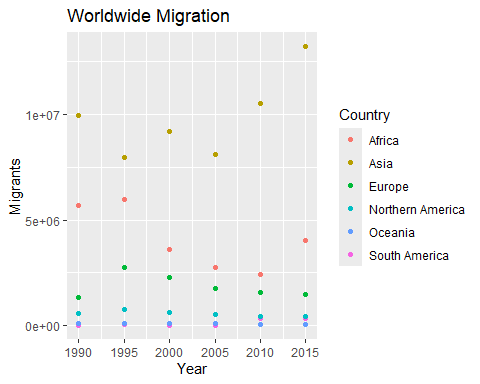
## # A tibble: 6 × 5  
## Country Country\_Code Type Year Cases  
## <chr> <dbl> <chr> <dbl> <dbl>  
## 1 WORLD 900 <NA> 1990 18836571  
## 2 WORLD 900 <NA> 1995 17853840  
## 3 WORLD 900 <NA> 2000 15827803  
## 4 WORLD 900 <NA> 2005 13276733  
## 5 WORLD 900 <NA> 2010 15370755  
## 6 WORLD 900 <NA> 2015 19577474

### Part 4 - Research Questions

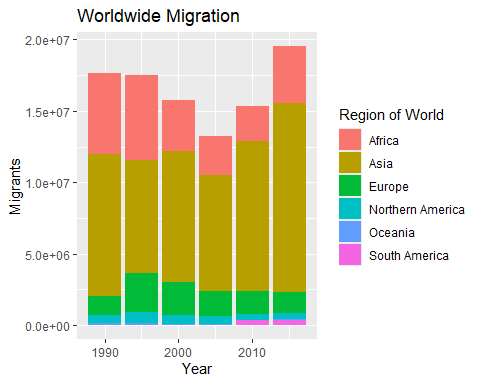
# Create subsets  
RegionalMigration <- filter(Migration2, Country %in% c("Africa", "Asia", "Europe", "Oceania", "Northern America", "South America"))  
Americas <- filter(Migration2, Country %in% c("Central America", "South America", "Northern America"))

### Worldwide Migration based on Regions

# Scatterplot   
ggplot(RegionalMigration, aes(x = Year, y = Cases, color = Country)) +  
 geom\_point() +  
 labs(title = "Worldwide Migration",  
 x = "Year",  
 y = "Migrants",  
 color = "Country")



# Bar Graph   
ggplot(RegionalMigration, aes(x = Year, y = Cases, fill = Country)) +  
 geom\_col() +  
 labs(title = "Worldwide Migration",  
 x = "Year",  
 y = "Migrants",  
 fill = "Region of World") +  
 scale\_fill\_discrete(name = "Region of World")

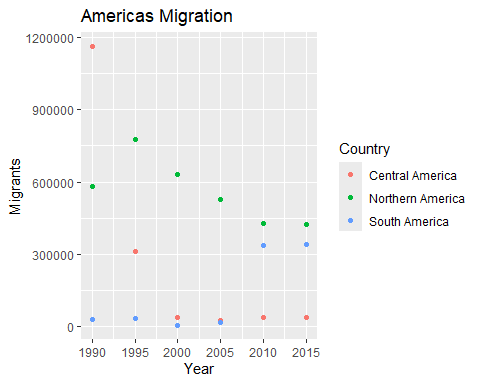


**a. Which region in the world had the highest number of migrants in the year 2005?**  
Asia  
**b. Over the years, which region consistently has the most migrants every 5 year span? Which has the second most?**  
Asia, Africa  
**c. What region has seen the fewest migrants over the years?**  
Oceania  
**d. Which plot was most useful in answering these questions and why?**  
While the scatterplot offered detailed insight into migration trends over time, the bar graph provided a concise summary of total migration counts across regions, making it equally valuable for understanding overall migration patterns. Both visualizations complemented each other well in providing a comprehensive view of the data.

### Americas Migration by Region

**a. In 1990, which region had the largest number of migrants for the Americas?**  
Central  
**b. Has this region continued to grow since 1990?**  
No  
**c. What trends do you notice happening in the Americas over the years?**  
Migration has decreased for Northern America and South America has increased between 2005 to 2010.  
**d. Specifically, has Northern America increased or decreased over the years?**  
Decreased  
**e. Which plot was most useful in answering these questions and why?**  
The scatterplot was more helpful because it makes it easier to see how migration trends change over time. With the bar graph, you have to compare the sizes of the blocks, which can be a bit trickier for understanding patterns.

# Scatterplot   
ggplot(Americas, aes(x = Year, y = Cases, color = Country)) +  
 geom\_point() +  
 labs(title = "Americas Migration",  
 x = "Year",  
 y = "Migrants",  
 color = "Country")



# Bar Graph  
ggplot(Americas, aes(x = Year, y = Cases, fill = Country)) +  
 geom\_col() +  
 labs(title = "Americas Migration",  
 x = "Year",  
 y = "Migrants",  
 fill = "Americas Region")

