# Module 5 - Assignment 1

## Angelo, Elizabeth

### Data Wrangling

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ 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`

view(UN\_migrant)

### Part 2 - Cleaning Data with dplyr

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`  
)  
head(UN\_migrant)

## # A tibble: 6 × 22  
## ...1 Country ...3 Country\_Code Type `1990` `1995` `2000` `2005` `2010`  
## <dbl> <chr> <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 WORLD <NA> 900 <NA> 1.88e7 1.79e7 1.58e7 1.33e7 1.54e7  
## 2 2 Developed r… (b) 901 <NA> 2.01e6 3.61e6 3.00e6 2.36e6 2.05e6  
## 3 3 Developing … (c) 902 <NA> 1.68e7 1.42e7 1.28e7 1.09e7 1.33e7  
## 4 4 Least devel… (d) 941 <NA> 5.05e6 5.16e6 3.05e6 2.36e6 1.96e6  
## 5 5 Less develo… <NA> 934 <NA> 1.18e7 9.08e6 9.78e6 8.55e6 1.14e7  
## 6 6 Sub-Saharan… (e) 947 <NA> 5.52e6 5.75e6 3.42e6 2.56e6 2.22e6  
## # ℹ 12 more variables: `2015` <dbl>, `1990...12` <chr>, `1995...13` <chr>,  
## # `2000...14` <chr>, `2005...15` <chr>, `2010...16` <chr>, `2015...17` <chr>,  
## # `1990-1995` <chr>, `1995-2000` <chr>, `2000-2005` <chr>, `2005-2010` <chr>,  
## # `2010-2015` <chr>

Migration <- UN\_migrant %>%  
 select(  
 Country,  
 Country\_Code,  
 Type,  
 "1990",  
 "1995",  
 "2000",  
 "2005",  
 "2010",  
 "2015"  
 )  
head(Migration)

## # A tibble: 6 × 9  
## Country Country\_Code Type `1990` `1995` `2000` `2005` `2010` `2015`  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 WORLD 900 <NA> 1.88e7 1.79e7 1.58e7 1.33e7 1.54e7 1.96e7  
## 2 Developed regions 901 <NA> 2.01e6 3.61e6 3.00e6 2.36e6 2.05e6 1.95e6  
## 3 Developing regio… 902 <NA> 1.68e7 1.42e7 1.28e7 1.09e7 1.33e7 1.76e7  
## 4 Least developed … 941 <NA> 5.05e6 5.16e6 3.05e6 2.36e6 1.96e6 3.44e6  
## 5 Less developed r… 934 <NA> 1.18e7 9.08e6 9.78e6 8.55e6 1.14e7 1.42e7  
## 6 Sub-Saharan Afri… 947 <NA> 5.52e6 5.75e6 3.42e6 2.56e6 2.22e6 3.64e6

Part 3 - Creating tidy data using tidyr

Migration2 <- Migration %>%  
 pivot\_longer(  
 cols = c("1990", "1995", "2000", "2005", "2010", "2015"),  
 names\_to = "year",  
 values\_to = "cases"  
 ) %>%  
mutate(year = as.numeric(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

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

## # A tibble: 6 × 5  
## Country Country\_Code Type year cases  
## <chr> <dbl> <chr> <dbl> <dbl>  
## 1 Africa 903 <NA> 1990 5687352  
## 2 Africa 903 <NA> 1995 5949953  
## 3 Africa 903 <NA> 2000 3609138  
## 4 Africa 903 <NA> 2005 2750644  
## 5 Africa 903 <NA> 2010 2413917  
## 6 Africa 903 <NA> 2015 4022363

head(Americas)

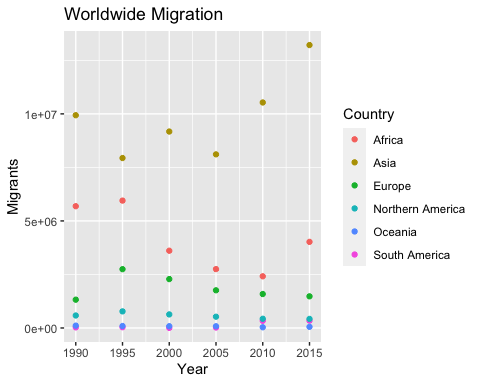
## # A tibble: 6 × 5  
## Country Country\_Code Type year cases  
## <chr> <dbl> <chr> <dbl> <dbl>  
## 1 Central America 916 <NA> 1990 1162216  
## 2 Central America 916 <NA> 1995 313977  
## 3 Central America 916 <NA> 2000 37295  
## 4 Central America 916 <NA> 2005 26083  
## 5 Central America 916 <NA> 2010 38361  
## 6 Central America 916 <NA> 2015 39982

### Worldwide Migration based on Regions

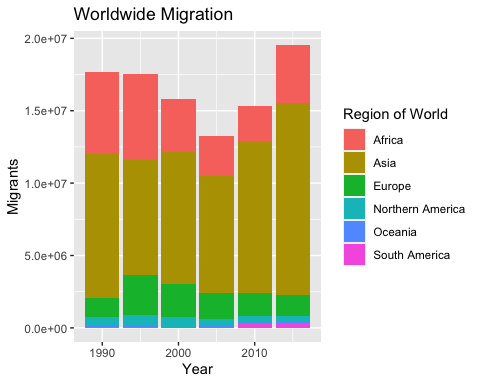
a. Which region in the world had the highest number of migrants in the year 2005?  
Asia.

b. Over the years, which region consistently has the most migrants every 5 year span? Which has the second most?  
Asia has the most migrants every five years, while Africa has the second most.  
  
c. What region has seen the fewest migrants over the years?  
Oceania  
  
d. Which plot was most useful in answering these questions and why?  
Both were useful. I found the bar graph useful to confirm the answer for question c, because the colors for each region were easy to read and differentiate. The scatterplot is useful because it is simple and clear to read, and made it easy to answer questions a and b.

library(ggplot2)  
scatterplot <- ggplot(RegionalMigration, aes(x = year, y = cases, color = Country)) +  
 geom\_point() +  
 labs(  
 title = "Worldwide Migration",  
 x = "Year",  
 y = "Migrants",  
 color = "Country"  
 )  
print(scatterplot)



bar\_graph <- ggplot(RegionalMigration, aes(x = year, y = cases, fill = Country)) +  
 geom\_bar(stat = "identity") +  
 labs(  
 title = "Worldwide Migration",  
 x = "Year",  
 y = "Migrants",  
 fill = "Region of World"  
 ) +  
 scale\_fill\_discrete(name = "Region of World")  
print(bar\_graph)

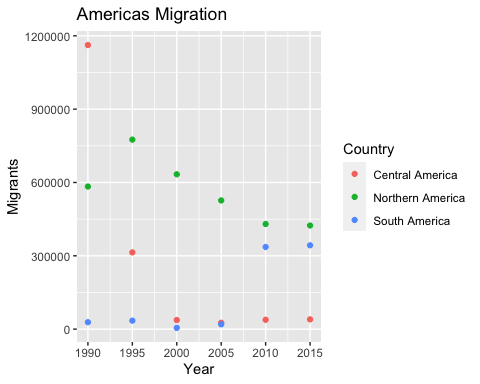


### 

### Americas Migration by Region

a. In 1990, which region had the largest number of migrants for the Americas?  
Central America  
  
b. Has this region continued to grow since 1990?  
Yes, but very little every five years.  
  
c. What trends do you notice happening in the Americas over the years?  
Migrants to the Americas seem to be decreasing over the years, and have not had equal to, or greater than 1,000,000 migrants since 1990-1995.  
  
d. Specifically, has Northern America increased or decreased over the years?  
Decreased.  
  
e. Which plot was most useful in answering these questions and why? I found the bar graph to be the most useful to answer questions a, b, and c. The bar graph was easy to read. The colors made it easy to see which region had the largest number of migrants in the Americas, and that it was not consistent growth. It also made it easier to notice migrant trends in the data. The scatterplot was the most useful to answer question d, because its data points showed a clear negative slope confirming a decrease in migrants over the years.

scatterplot\_americas <- ggplot(Americas, aes(x = year, y = cases, color = Country)) +  
 geom\_point() +  
 labs(  
 title = "Americas Migration",  
 x = "Year",  
 y = "Migrants",  
 color = "Country"  
 )  
print(scatterplot\_americas)



bar\_graph\_americas <- ggplot(Americas, aes(x = year, y = cases, fill = Country)) +  
 geom\_col() +  
 labs(  
 title = "Americas Migration",  
 x = "Year",  
 y = "Migrants",  
 fill = "Country"  
 ) +  
 scale\_fill\_discrete(name = "Americas Region")  
print(bar\_graph\_americas)

