# Module 5 - Assignment 1

## Black, Tyler

### 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 3.5.1 ✔ 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

UN\_migrant <- rename(UN\_migrant, "1990" = "1990...6", "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')  
  
Migration <- select(UN\_migrant,Country,Country\_Code,Type, '1990' ,'1995',  
 '2000','2005','2010','2015')

### Part 3 – Creating tidy data using tidyr

Migration2 <- pivot\_longer(Migration, cols = '1990':'2015', names\_to = 'Year', values\_to = 'Cases')  
Migration2$Year <- as.numeric(Migration2$Year)

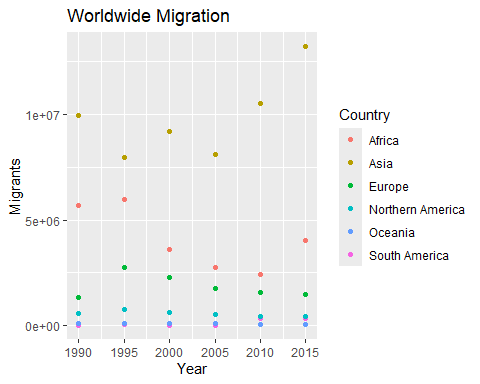
### Part 4 – Research Questions

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'))

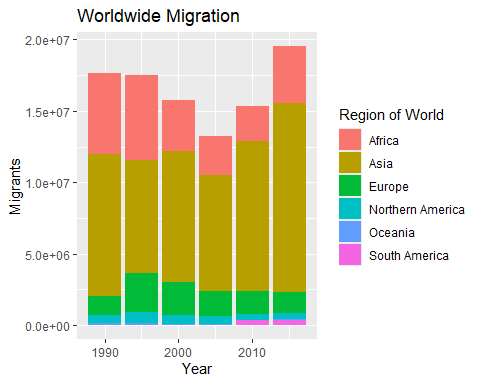
### Worldwide Migration based on Regions

1. Which region in the world had the highest number of migrants in the year 2005?  
   **Asia**
2. Over the years, which region consistently has the most migrants every 5 year span? Which has the second most?  
   **Asia and then Africa in second**
3. What region has seen the fewest migrants over the years?  
   **South America in 2000**
4. Which plot was most useful in answering these questions and why?  
   **I think the bar graph was the most helpful in answering these questions because of its simplicity and visualization**

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



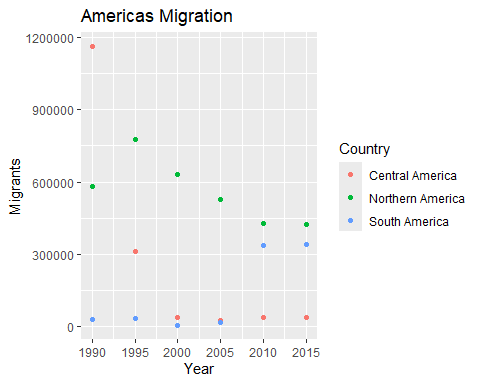
ggplot(RegionalMigration,mapping = aes(x=Year,y=Cases,fill = Country)) +  
 geom\_col() +   
 labs(title='Worldwide Migration', y= 'Migrants') +  
 scale\_fill\_discrete(name='Region of World')



### Americas Migration by Region

1. In 1990, which region had the largest number of migrants for the Americas?  
   **Central America**
2. Has this region continued to grow since 1990?  
   **Maybe internally, but the number of migrants has decreased over the years**
3. What trends do you notice happening in the Americas over the years?  
   **Migration has steadily decreased for Central and Northern America, but South America seems to have noticed a tick in migration in the most recent years**
4. Specifically, has Northern America increased or decreased over the years?  
   **North America has seen less and less migration over the years**
5. Which plot was most useful in answering these questions and why? **I think the scatter plot is the most useful in answering these questions because you can observe trends in the visualization of the data better than you can with the bar graph.**

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



ggplot(Americas,mapping = aes(x=Year,y=Cases,fill = Country)) +  
 geom\_col() +   
 labs(title='Americas Migration', y= 'Migrants') +  
 scale\_fill\_discrete(name='Americas Region')

