# Module 5 - Assignment 1

## Kiser, Stephen

### Data Wrangling

library(tidyverse)

## -- Attaching packages ------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.5  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts ---------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

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

## New names:  
## \* `` -> ...1  
## \* `` -> ...2  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5  
## \* ... and 12 more problems

View(UN\_migrant)

### Part 2 - Cleaning Data with dplyr

UN\_migrant <- rename(UN\_migrant, "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")  
  
head(Migration, 6)

## # A tibble: 6 x 9  
## Country Country\_Code Type `1990` `1995` `2000` `2005` `2010` `2015`  
## <chr> <dbl> <chr> <chr> <chr> <chr> <chr> <dbl> <dbl>  
## 1 WORLD 900 <NA> 188365~ 17853~ 15827~ 13276~ 1.54e7 1.96e7  
## 2 Developed regio~ 901 <NA> 2014564 36096~ 29972~ 23612~ 2.05e6 1.95e6  
## 3 Developing regi~ 902 <NA> 168220~ 14244~ 12830~ 10915~ 1.33e7 1.76e7  
## 4 Least developed~ 941 <NA> 5048391 51601~ 30474~ 23637~ 1.96e6 3.44e6  
## 5 Less developed ~ 934 <NA> 117736~ 90840~ 97830~ 85517~ 1.14e7 1.42e7  
## 6 Sub-Saharan Afr~ 947 <NA> 5516042 57478~ 34211~ 25550~ 2.22e6 3.64e6

### Part 3 - Creating tidy data using tidyr

Migration2 <- gather(Migration, "1990", "1995", "2000", "2005", "2010", "2015", key = "Year", value = "Cases")  
Migration2$Year <- as.integer(Migration2$Year)  
Migration2$Cases <- as.integer(Migration2$Cases)

## Warning: NAs introduced by coercion

head(Migration2, 6)

## # A tibble: 6 x 5  
## Country Country\_Code Type Year Cases  
## <chr> <dbl> <chr> <int> <int>  
## 1 WORLD 900 <NA> 1990 1.88e7  
## 2 Developed regions 901 <NA> 1990 2.01e6  
## 3 Developing regions 902 <NA> 1990 1.68e7  
## 4 Least developed countries 941 <NA> 1990 5.05e6  
## 5 Less developed regions excluding least devel~ 934 <NA> 1990 1.18e7  
## 6 Sub-Saharan Africa 947 <NA> 1990 5.52e6

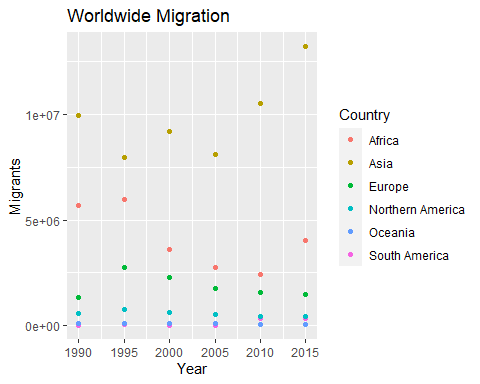
### Part 4 - Research Questions

RegionalMigration <- select(Migration2, Country, Country\_Code, Type, Year, Cases) %>%  
 filter(Country %in% c("Africa", "Asia", "Europe", "Oceania", "Northern America", "South America"))  
  
Americas <- select(Migration2, Country, Country\_Code, Type, Year, Cases) %>%  
 filter(Country %in% c("Central America", "South America", "Northern America"))

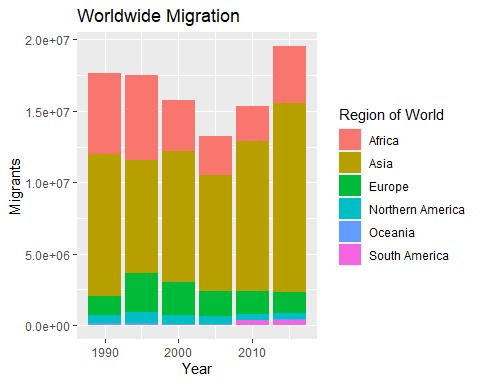
### Worldwide Migration based on Regions

1. Which region in the world had the highest number of migrants in the year 2005? Asia had the highest number of Migrants in 2005.
2. 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 5 year span and Africa has the second most.
3. What region has seen the fewest migrants over the years? South America has seen the fewest migrants over the years.
4. Which plot was most useful in answering these questions and why? The scatterplot graph was more useful because I was able to see the trends of each country better. It allowed me to see that the points for Asia were the highest every year. The bar graph would be better in see the totals of all migrants in the world, and where the majority of them were.

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



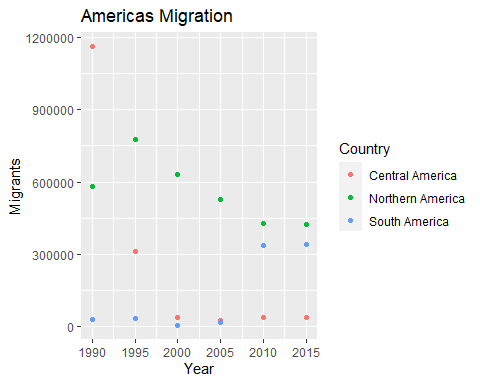
ggplot(RegionalMigration, aes(x=Year, y= Cases, fill = Country))+  
 geom\_col()+  
 labs(title = "Worldwide Migration", y="Migrants")+  
 scale\_fill\_discrete(name="Region of World")



### 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? No, after 1990 the number of migrants in Central America have decreased.
3. What trends do you notice happening in the Americas over the years? Central America started off high and then drastically decreased, Northern America had an increase in 1995 and then slowly decreased in migrants. Northern America started to plateau from 2010 to 2015. South America had almost no migrants from 1990 to 2010, and starting 2010 migrants increased and plateaued from 2010-2015.
4. Specifically, has Northern America increased or decreased over the years? Northern America has decreased over the years.
5. Which plot was most useful in answering these questions and why? The scatterplot was more useful in answering these questions because I was able to see the trend of the regions better than in the bar graph. It is hard to see the trend for Northern America from the bar graph, and was easier to see its trend in the scatterplot.

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



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

