# Module 5 - Assignment 1

## Graybill, Gil

### Data Wranglng

library("tidyverse")

## -- Attaching packages ------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- 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, "1990" = "1990...6",  
 "Country" = "...2",  
 "Country\_code" = "...4",  
 "Type" = "...5",  
 "1995" = "1995...7",  
 "2000" = "2000...8",  
 "2005" = "2005...9",  
 "2010" = "2010...10",  
 "2015" = "2015...11")

Migration <- UN\_migrant %>%  
select("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> 18836~ 17853~ 15827~ 13276~ 1.54e7 1.96e7  
## 2 Developed r~ 901 <NA> 20145~ 36096~ 29972~ 23612~ 2.05e6 1.95e6  
## 3 Developing ~ 902 <NA> 16822~ 14244~ 12830~ 10915~ 1.33e7 1.76e7  
## 4 Least devel~ 941 <NA> 50483~ 51601~ 30474~ 23637~ 1.96e6 3.44e6  
## 5 Less develo~ 934 <NA> 11773~ 90840~ 97830~ 85517~ 1.14e7 1.42e7  
## 6 Sub-Saharan~ 947 <NA> 55160~ 57478~ 34211~ 25550~ 2.22e6 3.64e6

### Part 3 Creating tidy data using tidyr

Migration2 <- Migration %>%  
 gather(`1990`, `1995`, `2000`, `2005`, `2010`, `2015`, key = "year", value = "cases")  
head(Migration2, 6)

## # A tibble: 6 x 5  
## Country Country\_code Type year cases   
## <chr> <dbl> <chr> <chr> <chr>   
## 1 WORLD 900 <NA> 1990 18836~  
## 2 Developed regions 901 <NA> 1990 20145~  
## 3 Developing regions 902 <NA> 1990 16822~  
## 4 Least developed countries 941 <NA> 1990 50483~  
## 5 Less developed regions excluding least d~ 934 <NA> 1990 11773~  
## 6 Sub-Saharan Africa 947 <NA> 1990 55160~

Migration2$year <- as.integer(Migration2$year)  
Migration2$cases <- as.integer(Migration2$cases)

## Warning: NAs introduced by coercion

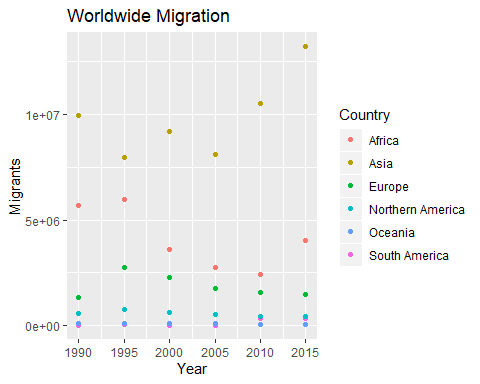
### Part 4 Research Questions

RegionalMigration <- Migration2 %>%  
 filter(Country == 'Africa' |   
 Country == 'Asia'|   
 Country == 'Europe'|   
 Country == 'Oceania'|   
 Country == 'Northern America'|   
 Country == 'South America')  
  
Americas <- Migration2 %>%  
 filter(Country == 'Central America' |   
 Country == 'South America'|   
 Country == 'Northern America')

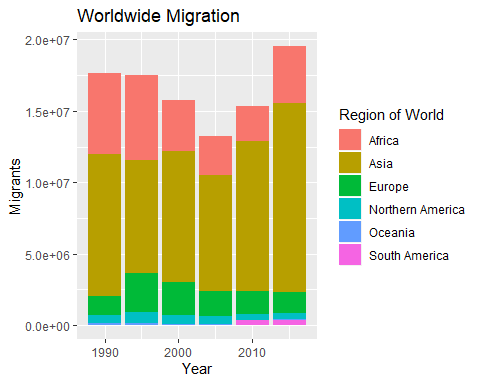
### Worldwide Migration based on Regions

1. Which region in the world had the highest number of migrants in the year 2005? ANSWER: Asia
2. Over the years, which region consistently has the most migrants every 5 year span? Which has the second most? ANSWER: Asia has the most, and Africa has the second most.
3. What region has seen the fewest migrants over the years? ANSWER: Oceania.
4. Which plot was most useful in answering these questions and why? It was easier to see on the scatterplot because of the colors and the low numbers for North America, South America, and Oceania. It’s slightly better if you use “scale\_fill\_brewer(name=”Region of World“, palette =”Dark2“)”

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

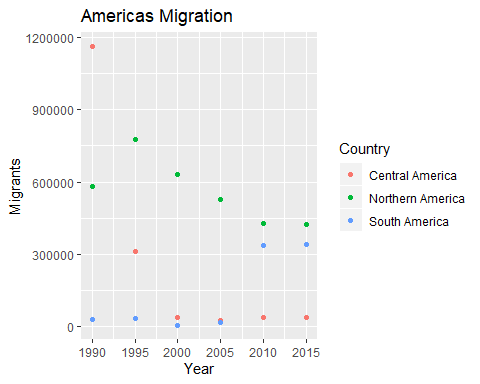


p1 <- ggplot(data = RegionalMigration, aes(x = year, y = cases, fill = Country))+  
 geom\_col()  
p1 +labs(title = "Worldwide Migration", y = "Migrants", x = "Year")+  
 scale\_fill\_discrete(name="Region of World")

 ### Americas Migration by Region a. In 1990, which region had the largest number of migrants for the Americas? ANSWER: Central America

1. Has this region continued to grow since 1990? ANSWER: No, it has been getting smaller
2. What trends do you notice happening in the Americas over the years? ANSWER: Migration is decreasing from Central America, increasing from South America.
3. Specifically, has Northern America increased or decreased over the years? ANSWER: Since 2000 it’s been almost constant
4. Which plot was most useful in answering these questions and why? ANSWER: It’s easier to get those answers from the bar chart

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



p1 <- ggplot(data = Americas, aes(x = year, y = cases, fill = Country))+  
 geom\_col()  
p1 +labs(title = "Worldwide Migration", y = "Migrants", x = "Year")+  
 scale\_fill\_discrete(name="Americas Region")

