# Module 6 - Assignment 1

## Broach, Stuart

### Data Wrangling

library("tidyverse")

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

## v ggplot2 3.1.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.7  
## v tidyr 0.8.2 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- 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)  
View(UN\_migrant)

### Part 2 - Cleaning Data with dplyr

UN\_migrant <- UN\_migrant %>%  
 rename(Country=X\_\_2, Country\_Code=X\_\_4, Type=X\_\_5)

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

## # 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 ~ 901 <NA> 20145~ 36096~ 29972~ 23612~ 2.05e6 1.95e6  
## 3 Developing~ 902 <NA> 16822~ 14244~ 12830~ 10915~ 1.33e7 1.76e7  
## 4 Least deve~ 941 <NA> 50483~ 51601~ 30474~ 23637~ 1.96e6 3.44e6  
## 5 Less devel~ 934 <NA> 11773~ 90840~ 97830~ 85517~ 1.14e7 1.42e7  
## 6 Sub-Sahara~ 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)

## # 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 ~ 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

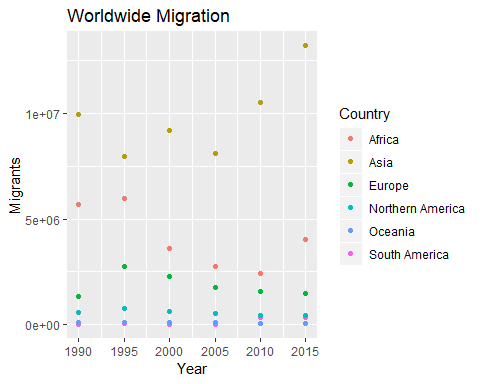
### Part 4 - Research Questions

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

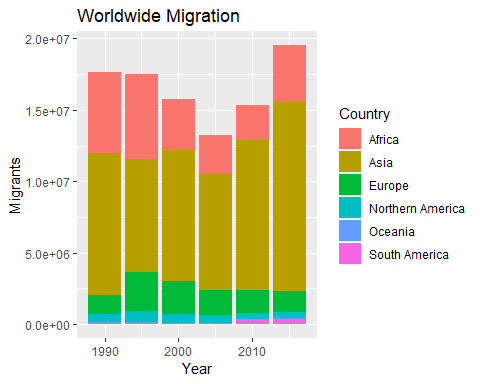
### Worldwide Migration based on Regions

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 followed by Africa. c.What region has seen the fewest migrants over the years? South America d.Which plot was most useful in answering these questions and why? The scaterplot because it didn’t bleed the colors together.

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



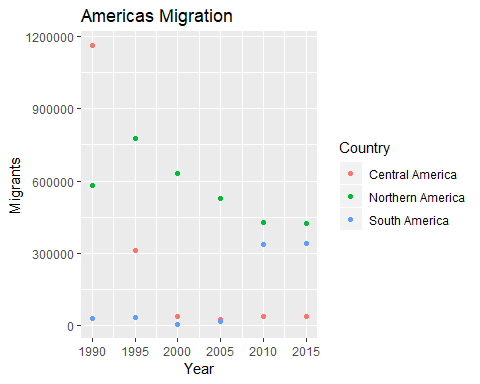
ggplot(RegionalMigration, aes(x = Year,y = Cases, fill = Country, scale\_fill\_discrete(name="Reion of World"))) +  
 geom\_col()+  
 labs(title = "Worldwide Migration",  
 x = "Year",  
 y = "Migrants")



### 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, it has decreased.
3. What trends do you notce happening in the Americas over the years? More people are migrating to Northern and South America.
4. Specifically, has Northern America increased or decreased over the years? Its stayed about the same but decreased some.
5. Which plot was most useful in answering these questions and why? the column chart because there was only 3 variables.

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



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

