# Module 5 - Assignment 1

## Bradley, Zackery

### Data Wrangling

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.2 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

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

library(dplyr)  
library(readxl)  
UN\_migrant <- read\_excel("C:/Users/bradl/OneDrive/Desktop/MODULE 2/Model 5/UN\_migrant.xlsx",   
 sheet = "Table 6", col\_types = c("numeric",   
 "text", "text", "numeric", "text",   
 "numeric", "numeric", "numeric",   
 "numeric", "numeric", "numeric",   
 "text", "text", "text", "text", "numeric",   
 "numeric", "text", "text", "text",   
 "text", "text"), skip = 15)

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in F40 / R40C6: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in G40 / R40C7: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in H40 / R40C8: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in I40 / R40C9: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in F179 / R179C6: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in G179 / R179C7: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in H179 / R179C8: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in I179 / R179C9: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in F220 / R220C6: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in G220 / R220C7: got '..'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in H220 / R220C8: got '..'

## New names:  
## \* `` -> ...1  
## \* `` -> ...2  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5  
## \* ...

#part 2 #==question 16 ### \*\*Cleaning Dta with dplyr

#== Question 17-18

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<- UN\_migrant[,c("Country", "Country\_Code", "Type", "1990", "1995", "2000", "2005", "2010", "2015")]

#==Part 3 #===Question 20-21

### Part 3- Creating tidy data using tidyr

#==question 22-24

Migration2 <- pivot\_longer(Migration,c("1990", "1995", "2000", "2005", "2010", "2015"),names\_to = "year", values\_to = "cases")  
head(Migration2)

## # A tibble: 6 x 5  
## Country Country\_Code Type year cases  
## <chr> <dbl> <chr> <chr> <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

#==question 23

Migration2$year<- as.numeric(Migration2$year)

#==question 25

####Research Questions

#question 26

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

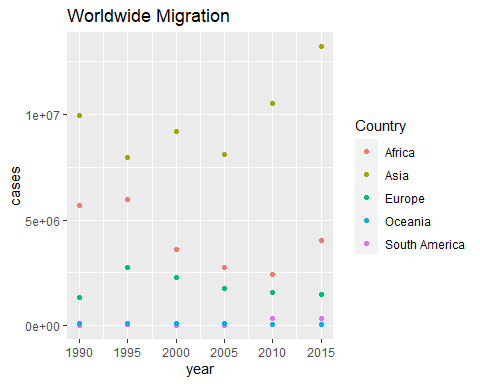
#==question 27-29

###Worldwide Migration Based on Regions

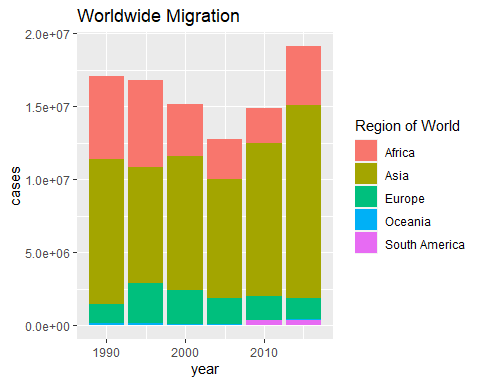
#a. Which region in the world had the highest number of migrants in the year 2005? Asia #b. Over the years, which region consistently has the most migrants every 5 year span? Which has the second most? Asia has the most, while Africa has the second most. #c. What region has seen the fewest migrants over the years? South America has seen the fewest migrants. #d. Which plot was most useful in answering these questions and why? I believe both plots were useful in answering questions, but the scatterplot was more useful. The scatterplot is more quickly understood and has more clear data than the bar chart.

#==Question 28

ggplot(data = RegionalMigration,mapping = aes(x = year, y = cases, color = Country))+  
ggtitle("Worldwide Migration")+  
geom\_point()



ggplot(data = RegionalMigration,mapping = aes(x = year, y = cases, fill = Country))+  
scale\_fill\_discrete(name="Region of World")+  
ggtitle("Worldwide Migration")+  
 geom\_col()

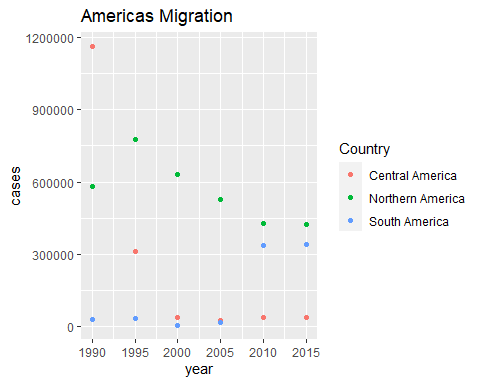


#===Question 30

###**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? This region has since declined in growth. #c. What trends do you notice happening in the Americas over the years? Central America is steadily decreasing in growth, while South America is generally increasing in growth. #d. Specifically, has Northern America increased or decreased over the years? North America increased in growth between 1990-1995, but has declined in growth since. #e. Which plot was most useful in answering these questions and why? The scatterplot was more useful in answering the questions because the data is more clean, concise, and easy to read.

ggplot(data = Americas, aes(x = year, y = cases, color = Country))+  
ggtitle("Americas Migration")+  
geom\_point()



ggplot(data = Americas, aes(x = year, y = cases, fill= Country))+  
ggtitle("Americas Migration")+  
scale\_fill\_discrete(name="Americas Region")+  
geom\_col()

