# Module5 - Assignment1

## Tate, Levi

### Data Wrangling

### Part 1: Importing the dataset using readxl

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.2 ✓ purrr 0.3.4  
## ✓ tibble 3.0.4 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ 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", 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 – Cleaning Data with dplyr

as\_tibble(UN\_migrant)

## # A tibble: 265 x 22  
## ...1 ...2 ...3 ...4 ...5 `1990...6` `1995...7` `2000...8` `2005...9`  
## <dbl> <chr> <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 1 WORLD <NA> 900 <NA> 18836571 17853840 15827803 13276733  
## 2 2 Deve… (b) 901 <NA> 2014564 3609670 2997256 2361229  
## 3 3 Deve… (c) 902 <NA> 16822007 14244170 12830547 10915504  
## 4 4 Leas… (d) 941 <NA> 5048391 5160131 3047488 2363782  
## 5 5 Less… <NA> 934 <NA> 11773616 9084039 9783059 8551722  
## 6 6 Sub-… (e) 947 <NA> 5516042 5747830 3421165 2555099  
## 7 7 Afri… <NA> 903 <NA> 5687352 5949953 3609138 2750644  
## 8 8 East… <NA> 910 <NA> 3168001 2046088 1641559 1419685  
## 9 9 Buru… <NA> 108 B R 267929 173017 27136 20681  
## 10 10 Como… <NA> 174 B 0 5 10 5  
## # … with 255 more rows, and 13 more variables: `2010...10` <dbl>,  
## # `2015...11` <dbl>, `1990...12` <chr>, `1995...13` <chr>, `2000...14` <chr>,  
## # `2005...15` <chr>, `2010...16` <dbl>, `2015...17` <dbl>, `1990-1995` <chr>,  
## # `1995-2000` <chr>, `2000-2005` <chr>, `2005-2010` <chr>, `2010-2015` <chr>

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")

### Part 3 – Creating tidy data using tidyr

Migration2 <- Migration %>%  
 pivot\_longer(c('1990', '1995', '2000', '2005', '2010', '2015'), names\_to = "year", values\_to = "cases")  
  
Migration2$year <- as.numeric(Migration2$year)

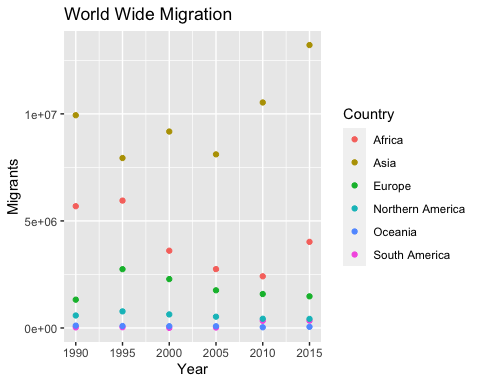
### Part 4 – Research Questions

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

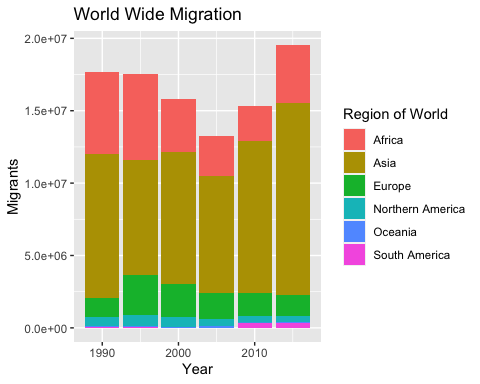
### Worldwide Migration based on Regions

1. Which region in the world had the highest number of migrants in the year 2005?  
   *Asia*
2. Over the years, which region consistently has the most migrants every 5 year span? Which has the second most?  
   *1. Asia | 2. Africa*
3. What region has seen the fewest migrants over the years?  
   *South America*
4. Which plot was most useful in answering these questions and why?  
   *The scatter plot was most useful because it is much easier to distinguish values.*

ggplot(RegionalMigration, mapping = aes(x = year, y = cases, color = Country)) +  
geom\_point() +  
labs(title = "World Wide Migration",  
 x = "Year",  
 y = "Migrants")



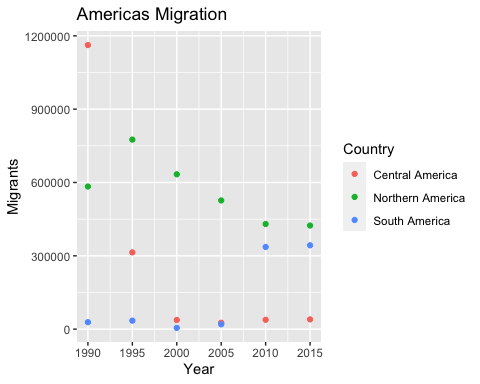
ggplot(RegionalMigration, mapping = aes(x = year, y = cases, fill = Country)) +  
geom\_col() +  
labs(title = "World Wide Migration",  
 x = "Year",  
 y = "Migrants") +  
scale\_fill\_discrete(name = "Region of World")



### 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*
3. What trends do you notice happening in the Americas over the years?  
   *Central America migration greatly decreases over the years, Northern America migration decreases as time goes on, and South America migration greatly increase after 2005.*
4. Specifically, has Northern America increased or decreased over the years?  
   *Northern America migration decreases over the years.*
5. Which plot was most useful in answering these questions and why?  
   *The scatter plot was most useful due to ease of viewing values.*

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



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

