title: “Visualisation in R, Week 5” author: “Your name here” output: word\_document

Experiment with the gapminder data to practice some of the new geoms we have learned. In this class, we focused on themes, colors, and labels. You should make sure all charts produced have all these elements. The output should be visually appealing. Please remember to comment on the findings.

Q1) Load packages needed, gapminder and explore/comment.

# Loads necessary packages  
library(ggplot2)   
# ggplot for data visualization  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

# dplyr for data manipulation  
library(gapminder)   
# gapminder is the dataset that we are working on   
  
head(gapminder)

## # A tibble: 6 × 6  
## country continent year lifeExp pop gdpPercap  
## <fct> <fct> <int> <dbl> <int> <dbl>  
## 1 Afghanistan Asia 1952 28.8 8425333 779.  
## 2 Afghanistan Asia 1957 30.3 9240934 821.  
## 3 Afghanistan Asia 1962 32.0 10267083 853.  
## 4 Afghanistan Asia 1967 34.0 11537966 836.  
## 5 Afghanistan Asia 1972 36.1 13079460 740.  
## 6 Afghanistan Asia 1977 38.4 14880372 786.

# displays the top rows of the data set  
summary(gapminder)

## country continent year lifeExp   
## Afghanistan: 12 Africa :624 Min. :1952 Min. :23.60   
## Albania : 12 Americas:300 1st Qu.:1966 1st Qu.:48.20   
## Algeria : 12 Asia :396 Median :1980 Median :60.71   
## Angola : 12 Europe :360 Mean :1980 Mean :59.47   
## Argentina : 12 Oceania : 24 3rd Qu.:1993 3rd Qu.:70.85   
## Australia : 12 Max. :2007 Max. :82.60   
## (Other) :1632   
## pop gdpPercap   
## Min. :6.001e+04 Min. : 241.2   
## 1st Qu.:2.794e+06 1st Qu.: 1202.1   
## Median :7.024e+06 Median : 3531.8   
## Mean :2.960e+07 Mean : 7215.3   
## 3rd Qu.:1.959e+07 3rd Qu.: 9325.5   
## Max. :1.319e+09 Max. :113523.1   
##

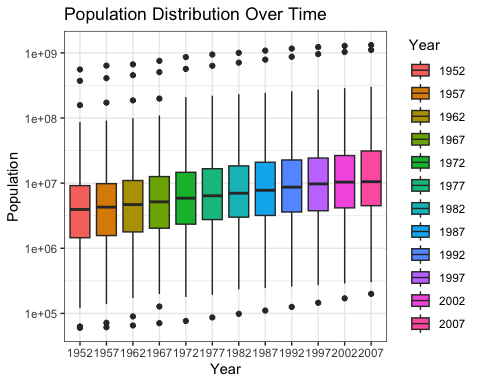
# prints the summary of the data set contains mean,medium,min,maximum  
str(gapminder)

## tibble [1,704 × 6] (S3: tbl\_df/tbl/data.frame)  
## $ country : Factor w/ 142 levels "Afghanistan",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ continent: Factor w/ 5 levels "Africa","Americas",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ year : int [1:1704] 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...  
## $ lifeExp : num [1:1704] 28.8 30.3 32 34 36.1 ...  
## $ pop : int [1:1704] 8425333 9240934 10267083 11537966 13079460 14880372 12881816 13867957 16317921 22227415 ...  
## $ gdpPercap: num [1:1704] 779 821 853 836 740 ...

# gives information about varaibles in the data set   
# The attributes in the data set are country, continent, year, lifeExp, pop and gdpPercap.  
#Life expectancy ranges from ~23 to ~82 years.  
#Population varies significantly across countries.  
#GDP per capita has a large variation, showing economic differences between countries.

Q2) Try examining population over time using a series of boxplots. (Hint: you may need to use the group aesthetic in the aes() call.)

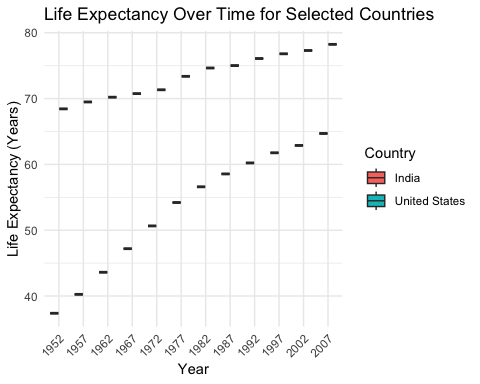
# Create a boxplot of population over time  
ggplot(gapminder, aes(x = factor(year), y = pop, group = year)) +   
 #sets ggplot with gapminder data, aes mapping factor of year to x-axis and population to y-axis, and groups by year  
 geom\_boxplot(aes(fill = factor(year))) +   
 # Boxplot to distribute the population over time  
 # Boxplot using different colors for each year  
 labs(title = "Population Distribution Over Time",  
 x = "Year",  
 y = "Population",  
 fill = "Year") +  
 theme\_bw() +   
 scale\_y\_log10()



# Log scale to y-axis to handle range of population values which vary greatly  
  
#Findings   
# From the plot we can observe that there is an increase in the population over time in 2007 compared to 1952.  
#The median population increases over time, meaning most countries experience growth.  
#There are some extreme outliers, likely countries with very large populations (e.g., China, India).  
#Africa's population is increasing steadily, but variation remains high.  
#Oceania has the smallest populations, with less variation across years.

Q3) Try examining life expectancy over time using a series of boxplots.

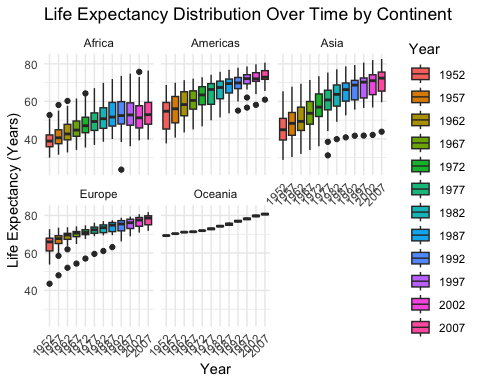
# Select a few representative countries for comparison  
selected\_countries <- c("United States", "India")  
# Filter the dataset to include only the selected countries  
gapminder\_filtered <- gapminder %>%   
 filter(country %in% selected\_countries)  
# Create a boxplot for life expectancy over time by country  
ggplot(gapminder\_filtered, aes(x = factor(year), y = lifeExp, fill = country)) +   
 # sets ggplot with gapminder data, aes mapping factor of year to x-axis and life expectancy to y-axis  
 geom\_boxplot() +   
 # Creates the boxplot for life expectancy over time  
 theme\_minimal() +   
 labs(title = "Life Expectancy Over Time for Selected Countries",   
 x = "Year", # Label for the x-axis (time in years)  
 y = "Life Expectancy (Years)", # Label for the y-axis  
 fill = "Country") + # Legend label for colors  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# Rotates x-axis labels for readability  
  
#Findings   
#The United States shows a higher median life expectancy across all years compared to India.  
#The range of values in the U.S. is narrower, indicating less variation across different data points.  
#India Shows a Significant Increase in Life Expectancy  
#In 1952, India's life expectancy was much lower compared to the U.S. (~37 years vs. ~68 years).  
#Over time, India's median life expectancy increased steadily.

Q4) Can you facet this boxplot by continent? If so please do and explain how this differs.

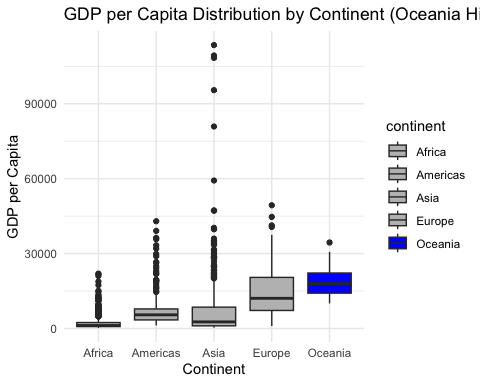
# Create a faceted boxplot by continent  
ggplot(gapminder, aes(x = factor(year), y = lifeExp, group = year)) +  
 # sets ggplot with gapminder data, aes mapping factor of year to x-axis and life expectancy to y-axis, and groups by year  
 geom\_boxplot(aes(fill = factor(year))) +  
 # creating a boxplot using diffrent colors   
 facet\_wrap(~continent) + # Facet by continent  
 theme\_minimal() +  
 labs(title = "Life Expectancy Distribution Over Time by Continent",  
 x = "Year",  
 y = "Life Expectancy (Years)",  
 fill = "Year") +  
 # adding the labels to x and y axes along with title  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# setting the element text abgle to 45 so that we can clearly indetifies the. text in the theme   
  
#Findings  
#Life expectancy has increased over time across all countries.  
#The median life expectancy rises in each subsequent year.  
#In 1952, the boxplots show a wide spread, meaning some countries had very low life expectancy (~30 years), while others had much higher values (~70+ years).  
#By 2007, the spread has narrowed, indicating that fewer countries have extremely low life expectancy.  
#The spread (IQR) narrows over time, indicating that global health improvements have led to less variation between countries.

Q5) Make a box plot of gdpPercap per continent, we are interested in “Oceania”, please make this one stand out like we did in class.

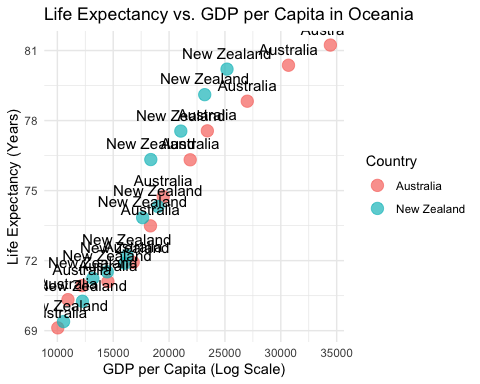
# Boxplot is created for gdpPercap per continent and highlighting Oceania  
ggplot(gapminder, aes(x = continent, y = gdpPercap, fill = continent)) +  
 # ggplot with gapminder data, aes maps x-axis to continent and y-axis to GDP per Capita, and fill color according to continent  
 geom\_boxplot() + # creates the boxplot  
 scale\_fill\_manual(values = c("grey", "grey", "grey", "grey", "blue")) +  
 #manually setting the fill colors, with four continents in grey and one in red to ensure Oceania is highlighted  
 labs(title = "GDP per Capita Distribution by Continent (Oceania Highlighted)",  
 x = "Continent",  
 y = "GDP per Capita")+  
 theme\_minimal() # setting minimal theme



# Findings  
#Oceania has the highest median GDP per capita, with outliers like Australia and New Zealand clearly standing out.  
#The BLUE color highlights the prominence of Oceania compared to the other continents.

Q6) Let’s look more into Oceania now. Make a scatterplot with arranged labels. This graph should be visually appealing and not be “busy”. You may choose the two variables used. Explain why they are important to compare, they should make sense.

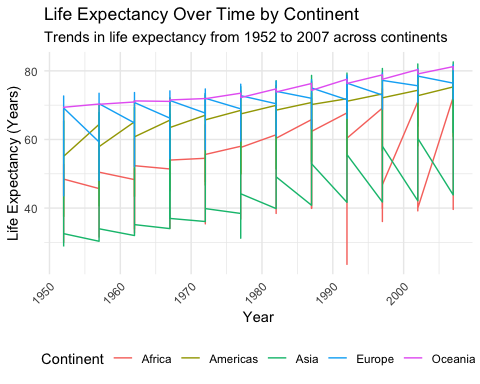
# Filter the gapminder data for Oceania  
oceania\_data <- gapminder %>%   
 filter(continent == "Oceania")  
# Create a scatterplot with arranged labels  
ggplot(oceania\_data, aes(x = gdpPercap, y = lifeExp, label = country)) +  
# ggplot with Oceania data, aes maps x-axis to GDP per Capita and y-axis to Life Expectancy , and labels with country  
 geom\_point(aes(color = country), size = 4, alpha = 0.7) +   
 # Points for each country, color-coded by country  
 geom\_text(aes(label = country), vjust = -1, hjust = 0.5, size = 4, color = "black") +   
 # Adding labels for each country  
 labs(title = "Life Expectancy vs. GDP per Capita in Oceania",  
 x = "GDP per Capita (Log Scale)",  
 y = "Life Expectancy (Years)",  
 color = "Country") +  
 # labeling x and y axes and adding title  
 # Color legend for countries  
 theme\_minimal() # sets minimal theme



#Findings  
#Both countries have high GDP per capita and high life expectancy, making them standout in the scatterplot.  
#These two countries show the typical correlation where higher GDP correlates with longer life expectancy, which is expected in developed nations.  
#There are no countries in Oceania with low GDP and high life expectancy, which could indicate that economic development plays a significant role in life expectancy in this region

Q7) Plot lifeExp over years, grouped by continent. Please choose a theme from class and have a title, sub title, some comments on the graph, and a title for the legend, and place the legend on the bottom. Please note you should think about ‘group()’.

# Create the plot for life expectancy over years, grouped by continent  
ggplot(gapminder, aes(x = year, y = lifeExp, color = continent, group = continent)) +   
 # ggplot with gapminder data, aes maps year to x-axis and life expectancy to y-axis, using continent for color and groups by continent, year  
 geom\_line() + # Draw lines to show life expectancy trends  
 #creates lineplot for each continent  
 theme\_minimal() + # Clean theme  
 labs(  
 title = "Life Expectancy Over Time by Continent", # Main title  
 subtitle = "Trends in life expectancy from 1952 to 2007 across continents",   
 # adding labels for x-axis and the y-axis  
 x = "Year",   
 y = "Life Expectancy (Years)",  
 color = "Continent" # Title for the legend  
 ) +   
 theme(  
 axis.text.x = element\_text(angle = 45, hjust = 1),   
 # Rotate x-axis labels for better readability  
 legend.position = "bottom"   
 # Place the legend at the bottom  
 )



#Findings  
#The plot uses lines to show the trend of life expectancy over time for each continent.  
#This highlights the differences and trends in life expectancy between continents.  
#The legend is placed at the bottom of the plot, as requested, for clarity.  
#Oceania (Australia & New Zealand) has the highest life expectancy, with steady increases throughout the years.  
#Europe and Americas show relatively consistent life expectancy with slight increases, suggesting well-established healthcare and living conditions.  
#Asia and Africa have a wider range of life expectancy values, with significant increases in countries like China and India over time. However, Africa shows slower improvements overall, possibly due to ongoing health and socio-economic challenges.