**Malik Hassan Qayyum**

Take the task from Exam 1

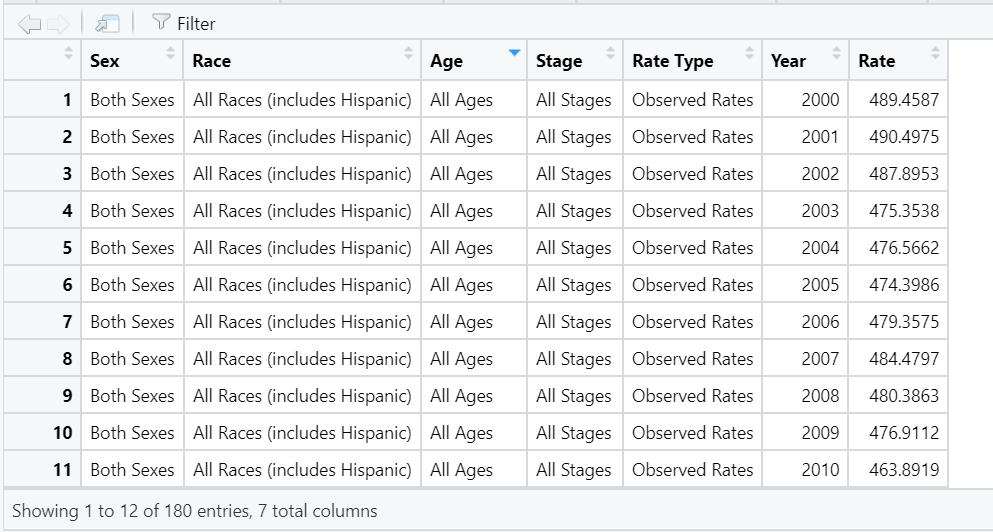
251/502

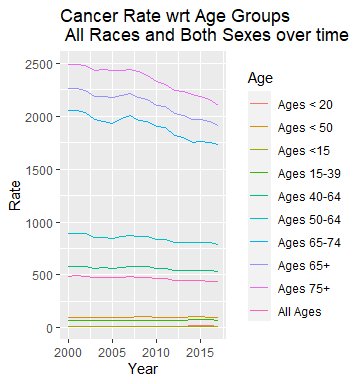
For this assignment, I want you to become comfortable manipulating dataframes and being able to plot the same data represented in two different formats using variations of ggplot (colour and geom\_line).

Long to wide AND wide to long data transformation exercise

The dataset is currently in a long format.

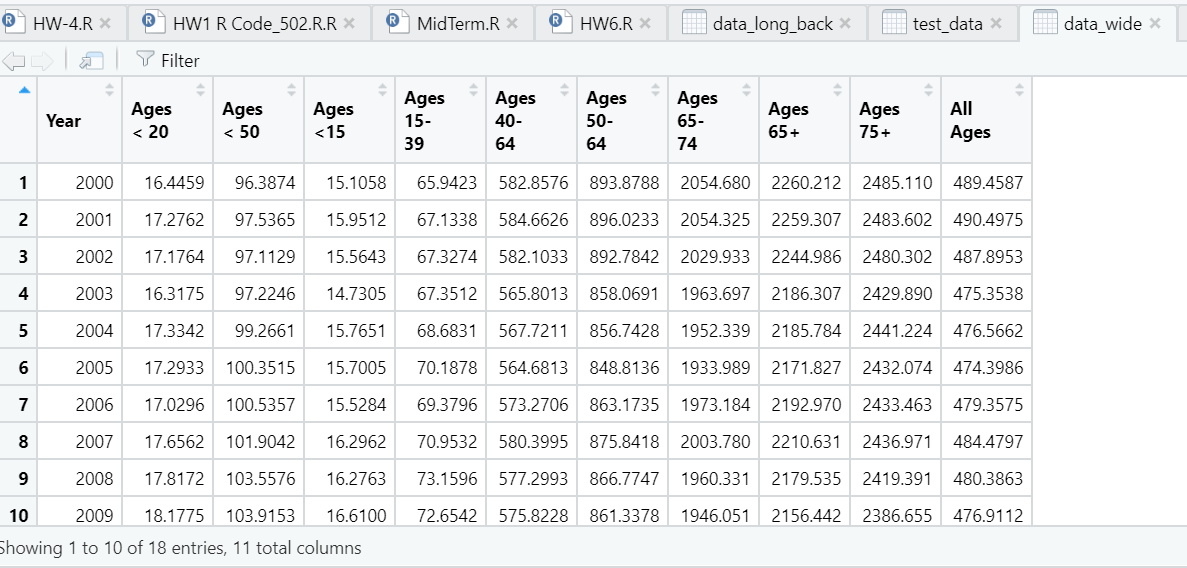
1. Construct **a time plot** of ‘Both Sexes’, ‘All Races (Includes Hispanics)’ across all and individual ages done on the exam. Paste a brief snapshot of the dataframe and your resulting graph and code using ggplot.

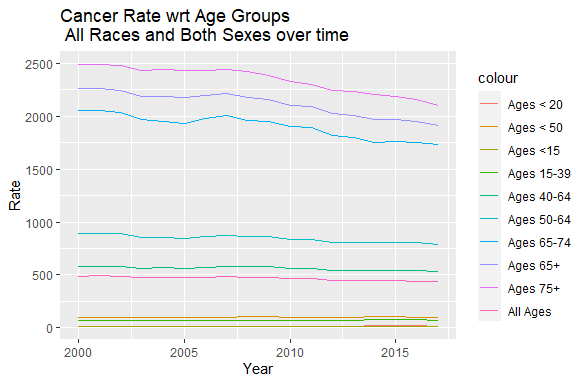




*All code is at the bottom.*

1. Convert the dataframe to wide format and do the same plot above using a different ggplot configuration. Supply a snapshot of your dataframe, graph, and code.





*All code is at the bottom.*

1. Manipulate the dataframe back to the long format, supply your code

*All code is at the bottom.*

Hints\*

<http://www.cookbook-r.com/Manipulating_data/Converting_data_between_wide_and_long_format/>

For data transformations: I like to use melt and cast, or transpose. You can also subset by row and column indexes and then cbind or rbind depending.

For ggplot: mentioned above, but also use ggplot with additional geom\_line() function for each additional age line.

(From this point forward, it need not be said that you default to an x-axis and y-axis labels, along with a title, and legend)

**Code:**

library(cowplot)

library(ggplot2)

library(scales)

library(reshape2)

library(xlsx)

#############################################################################

data <-read.csv("C:/Users/Malik/Documents/GitHub/Data-Visualization-Data502/exam1part2/exam 1 dataset.csv",

                na.strings = c("-"))

# Structure of data

str(data)

# Changing Column Names

colnames(data)

colnames(data)<- c("Sex", "Race", "Age", "Stage", "Rate Type", "Year", "Rate")

# 1.  Construct a time plot of ‘Both Sexes’, ‘All Races (Includes Hispanics)’

# across all and individual ages done on the exam.

dat = subset(data,data$Sex ==c("Both Sexes"))

dat = subset(dat,dat$Race == "All Races (includes Hispanic)")

#####################################################################################

# plotting long format data

#####################################################################################

# Time Plot

ggplot(dat, aes(x=Year, y=Rate,color=Age)) + geom\_line() +

  xlab("Year") + ylab("Rate") +

  ggtitle("Cancer Rate wrt Age Groups  \n All Races and Both Sexes over time")

#####################################################################################

# Converting long to wide format data

#####################################################################################

data\_wide <- dcast(dat, Year ~ Age, value.var="Rate")

colnames(data\_wide)

#####################################################################################

# Plotting wide format data

#####################################################################################

# get() gets the value name of the variable. As my varibales has spaces I am not able to use it directly that's why i am using get()

ggplot(data\_wide, aes(Year)) +

  geom\_line(aes(y = get("Ages < 20"), colour = "Ages < 20")) +

  geom\_line(aes(y = get("Ages < 50"), colour = "Ages < 50")) +

  geom\_line(aes(y = get("Ages <15") , colour ="Ages <15")) +

  geom\_line(aes(y = get("Ages 15-39"), colour = "Ages 15-39")) +

  geom\_line(aes(y = get('Ages 40-64'), colour ="Ages 40-64")) +

  geom\_line(aes(y = get("Ages 50-64") , colour ="Ages 50-64")) +

  geom\_line(aes(y = get("Ages 65-74"), colour ="Ages 65-74")) +

  geom\_line(aes(y = get("Ages 65+"), colour ="Ages 65+")) +

geom\_line(aes(y = get("Ages 75+"), colour ="Ages 75+")) +

  geom\_line(aes(y = get("All Ages") , colour = "All Ages")) +

  xlab("Year") + ylab("Rate") +

  ggtitle("Cancer Rate wrt Age Groups  \n All Races and Both Sexes over time")

#####################################################################################

# Converting wide to long format data

#####################################################################################

data\_long\_back = melt(data\_wide, id.vars="Year" ,

                      variable.name="Age",

                      value.name="Rate")

#####################################################################################

# References

# https://stackoverflow.com/questions/3777174/plotting-two-variables-as-lines-using-ggplot2-on-the-same-graph

# http://www.cookbook-r.com/Manipulating\_data/Converting\_data\_between\_wide\_and\_long\_format/

#####################################################################################

**End**