Problem Set 2

Name:

About this format: This problem set is saved as an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code. You can click "knit" at the top to create a pdf (which is what you will turn in). You can write text (just like in Word) in this document to answer the questions. You can then use the "code chunks" to write your R code. When you execute code within the notebook, the results appear beneath the code. To add a new R chunk click the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I* (you shouldn't need to to this since this document is set up with code chunks for each question, but just in case!).

To submit: Please upload the final .pdf file on Canvas

I. CO2 Trends

First read in the data and get it ready to use You can find the data and read more about it here: https://climate.nasa.gov/vital-signs/carbon-dioxide/ In particular, the metadata is at the top of the txt file if you click on the Download Data button.

To get you started here's the code to read in the data and give it better column names (you should be able to just run this - no need to modify it):

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

1. Read about the data
    Why are there three different columns for CO2? Where do the data come from? What does number-Days refer to? What does -9.99 in the stdev column mean? [1 point]

This is my answer.

2. Visualize the data: How do atmospheric CO2 concentrations change over the time series?

a. Describe your process: What category of graph will you use? Does the underlying data differ if you use the interpolated or trend columns? [1 point]

b. Graph it, with custom axis labels, a title, and change the color and theme: [3 points]

```r
str(CO2data)
```

```
## Classes 'data.table' and 'data.frame':   737 obs. of  8 variables:
##  $ year             : int  1959 1959 1959 1959 1960 1960 1960 1960 1960 1960 ...
##  $ month            : int  9 10 11 12 1 2 3 4 5 6 ...
##  $ decimalDate      : num  1960 1960 1960 1960 1960 ...
##  $ meanCO2          : num  314 313 315 316 316 ...
##  $ deseasonalizedCO2: num  317 316 317 316 316 ...
##  $ numberDays       : int  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
##  $ stdDays          : num  -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 -9.99 ...
##  $ uncMeanCO2       : num  -0.99 -0.99 -0.99 -0.99 -0.99 -0.99 -0.99 -0.99 -0.99 -0.99 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```r
min(CO2data$meanCO2)
```

```
## [1] 313.33
```

3. Visualize the data: Which months are the CO2 values at the maximum? Minimum? It may help to plot it below to answer this. Does this make sense (i.e., what might explain this)? [1 point]

a. Describe your process: What category of graph will you use? [1 points]

b. Graph it, with custom axis labels, a title, and a new theme (Note- it may be help to change month to a factor using the as.factor() function): [3 points]

II. Temperature trends

First we'll read in the data and get it ready to use You can find the data and read about it here: https://climate.nasa.gov/vital-signs/global-temperature/

```r
tempdata <- fread('http://climate.nasa.gov/system/internal_resources/details/original/647_Global_Tempera
```

```
## Warning in fread("http://climate.nasa.gov/system/internal_resources/details/
## original/647_Global_Temperature_Data_File.txt"): Detected 1 column names but the
## data has 3 columns (i.e. invalid file). Added 2 extra default column names at
## the end.
```

1. Read about the data What information is in each of the three columns? Where do the data come from? Using the names() function to give tempdata descriptive column headings [1 points]

2. Visualize the data

a. Recreate the graph on the website using all three columns (in terms of basic properties - no need to try to get the theme exactly the same). [2.5 points]

b. Recreate the graph on the website using two columns and a statistical transformation (Note - it won't look exactly the same due to differences in the default degree of smoothing) [2.5 points]

III. Ice sheets

Have a look at the following data, and read it into R Either use fread() in library(data.table) to read it from the web, or read.csv()

Data description: http://climate.nasa.gov/vital-signs/land-ice/ Raw data file: http://climate.nasa.gov/system/internal_resources/details/original/499_GRN_ANT_mass_changes.csv

```r
# Run the following to read the .csv from the web and rename columns
icedat0 <- fread('http://climate.nasa.gov/system/internal_resources/details/original/499_GRN_ANT_mass_cl

icedat <- icedat0[-c(1:10),]
names(icedat) = c("year_decimal", "Greenlandmass_Gt", "Antarcticamass_Gt")
icedat$year_decimal <- as.numeric(icedat$year_decimal)
icedat$Greenlandmass_Gt <- as.numeric(icedat$Greenlandmass_Gt)
icedat$Antarcticamass_Gt <- as.numeric(icedat$Antarcticamass_Gt)
```

1. Describe the data set: What are the columns and units? Where do the numbers come from? [1 point]

2. Plot a time series of the with one line for the Antarctic mass and one for the Greenland mass, with a separate line color for each. Describe the trends you observe. [3 points]