

Problem Set 1

Due 24 January

What to submit and how

Answer the questions below to the best of your ability. Write clearly, and format your tables and visuals appropriately. You must use R **Markdown** to compose and compile your work. For full credit, **echo** all code chunks. Submit your work in hard copy at the beginning of class.

Questions

Use the Gapminder data to extend analyses in sections 5.3-5.4 of Dr. Jenny Bryan's Stat545 book. Prof. Bryan relies on an abridged version of the dataset, `gapminder`. Your analysis must use the full `gapminder_unfiltered` data. To access the data, install and load the `gapminder` package. I recommend that you then assign the data to a new object (e.g., `df = gapminder_unfiltered`).

1. In a few sentences, describe the unfiltered Gapminder data frame. Be sure to note the number of observations, available years of data, and number of variables.
2. Plot life expectancy (`lifeExp`) by year.
3. Present a table/tabulation of the number of observations in each continent.
4. Recreate the plot at the end of section 5.5 using the unfiltered data. The plot shows the relationship between life expectancy and GDP per capita (`gdpPercap`) by continent. Present only the final plot (i.e., clean the code to omit all the intermediate plots).
5. Correct the code below! Wickham et al. (Chapter 4) offer a basic style guide for coding. Use their guide to clean up the awful, but functional, code below. *Note:* be sure to set `eval = FALSE` for this code chunk.

```
flights|>filter(dest=="IAH")|>group_by(year,month,day)|>summarize(n=n(),
delay=mean(arr_delay,na.rm=TRUE))|>filter(n>10)

flights|>filter(carrier=="UA",dest%in%c("IAH","HOU"),sched_dep_time>
0900,sched_arr_time<2000)|>group_by(flight)|>summarize(delay=mean(
arr_delay,na.rm=TRUE),cancelled=sum(is.na(arr_delay)),n=n())|>filter(n>10)
```