Noah Abrigo

Dr. Halil Bisgin

CSC 302

8 March 2022

<div align="center">Homework 4</div>

GitHub Link: https://github.com/noahabrigo/HW4

Part 1:

```python
#please use this cell to read and select your data

df = pd.read_csv("/content/drive/MyDrive/DATA/olympic_medals.csv")
df = df[df['Year'] >= 2000]
df
```

	Gender	Event	Location	Year	Medal	Name	Nationality	Result
0	M	10000M Men	Rio	2016	G	Mohamed FARAH	GBR	25:05.17
1	M	10000M Men	Rio	2016	S	Paul Kipngetich TANUI	KEN	27:05.64
2	M	10000M Men	Rio	2016	B	Tamirat TOLA	ETH	27:06.26
3	M	10000M Men	Beijing	2008	G	Kenenisa BEKELE	ETH	27:01.17
4	M	10000M Men	Beijing	2008	S	Sileshi SIHINE	ETH	27:02.77
...	...	...	...	...	...	...	...	...
2386	W	Triple Jump Women	London	2012	S	Caterine IBARGUEN	COL	14.8
2387	W	Triple Jump Women	London	2012	B	Olga SALADUKHA	UKR	14.79
2388	W	Triple Jump Women	Athens	2004	G	Francoise MBANGO ETONE	CMR	15.3
2389	W	Triple Jump Women	Athens	2004	S	Hrysopiyi DEVETZI	GRE	15.25
2390	W	Triple Jump Women	Athens	2004	B	Tatyana LEBEDEVA	RUS	15.14

```
#Please use this cell to create your your figure. Please use Year column to color your graph.
import plotly.express as px
fig = px.parallel_categories(df, dimensions=['Year', 'Medal', 'Gender'],
            color="Year", color_continuous_scale=px.colors.sequential.Inferno)
fig.show()
```
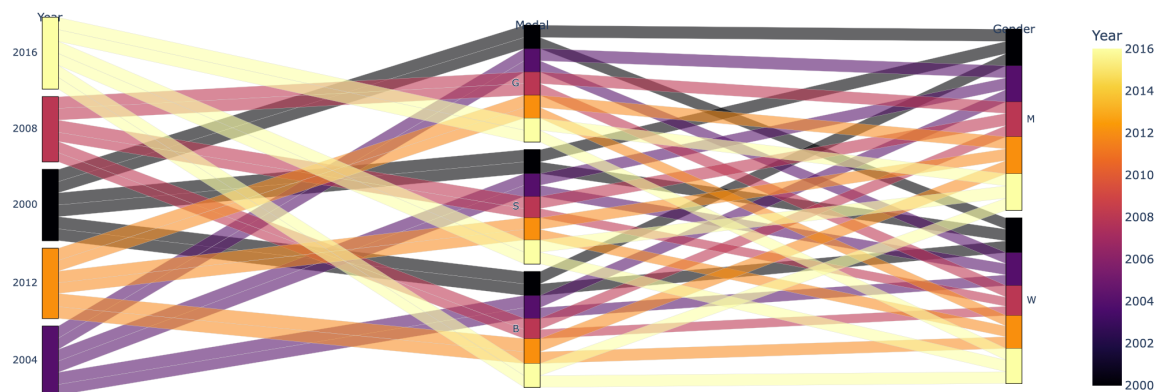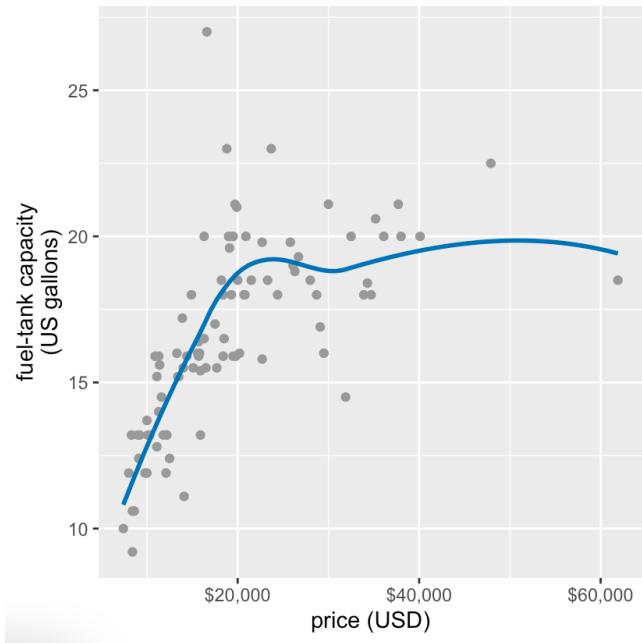


```
#You can use this cell to write your code. It is doable at most 4 lines of code.
df = pd.read_csv("/content/drive/MyDrive/DATA/2016elections.csv")
df = df[(df['st'] == 'AR') | (df['st'] == 'CA') | (df['st'] == 'MI') | (df['st'] == 'WI')]
df = df[['pct_clinton','pct_trump','pct_johnson','pct_other']]
df.plot(kind='bar', stacked=True)
plt.xticks([0, 1, 2, 3], ['AR','CA','MI','WI'])
```

```
([<matplotlib.axis.XTick at 0x7f0b4eae7510>,
  <matplotlib.axis.XTick at 0x7f0b4eae7310>,
  <matplotlib.axis.XTick at 0x7f0b4ea84550>,
  <matplotlib.axis.XTick at 0x7f0b4ea4fc50>],
 [Text(0, 0, 'AR'), Text(0, 0, 'CA'), Text(0, 0, 'MI'), Text(0, 0, 'WI')])
```
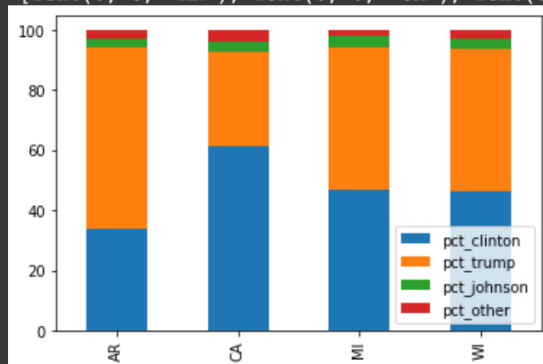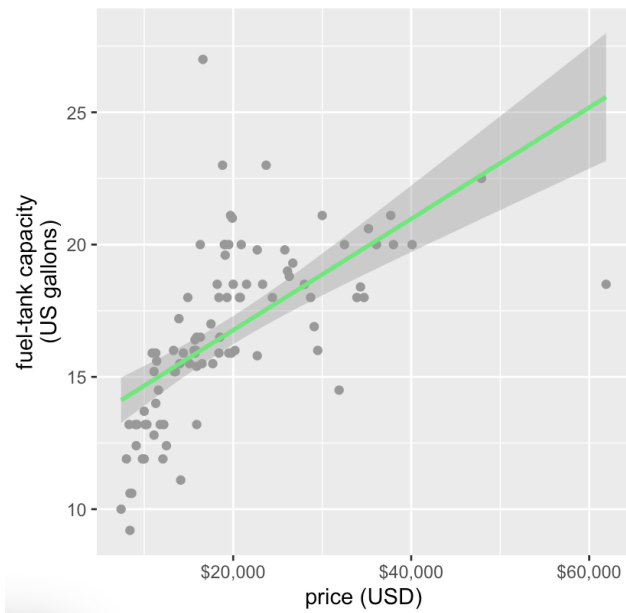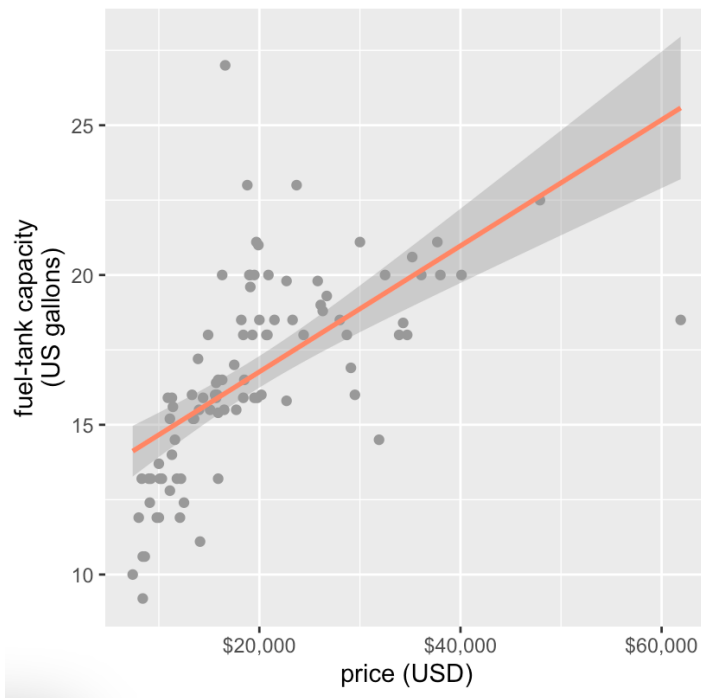
Part 2:

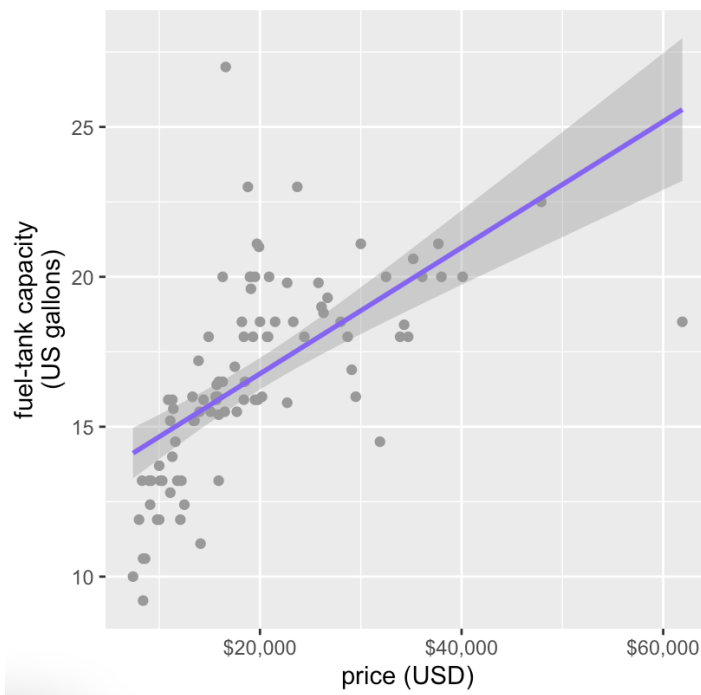First starting plot:

LM Smoothing

GLM Smoothing

GAM Smoothing

```
library(ggplot2)
cars93 <- MASS::Cars93
ggplot(cars93, aes(x = Price, y = Fuel.tank.capacity)) +
  geom_point(color = "grey60") +
  geom_smooth(se = TRUE, method = "gam", formula = y ~ x, color = "#7c6bea") +
  scale_x_continuous(
    name = "price (USD)",
    breaks = c(20, 40, 60),
    labels = c("$20,000", "$40,000", "$60,000")
  )+
  scale_y_continuous(name = "fuel-tank capacity\n(US gallons)")+
  ggtitle("GAM Smoothing")+
  theme(plot.title = element_text(size=14, color="#7c6bea"))
```

Part 3:
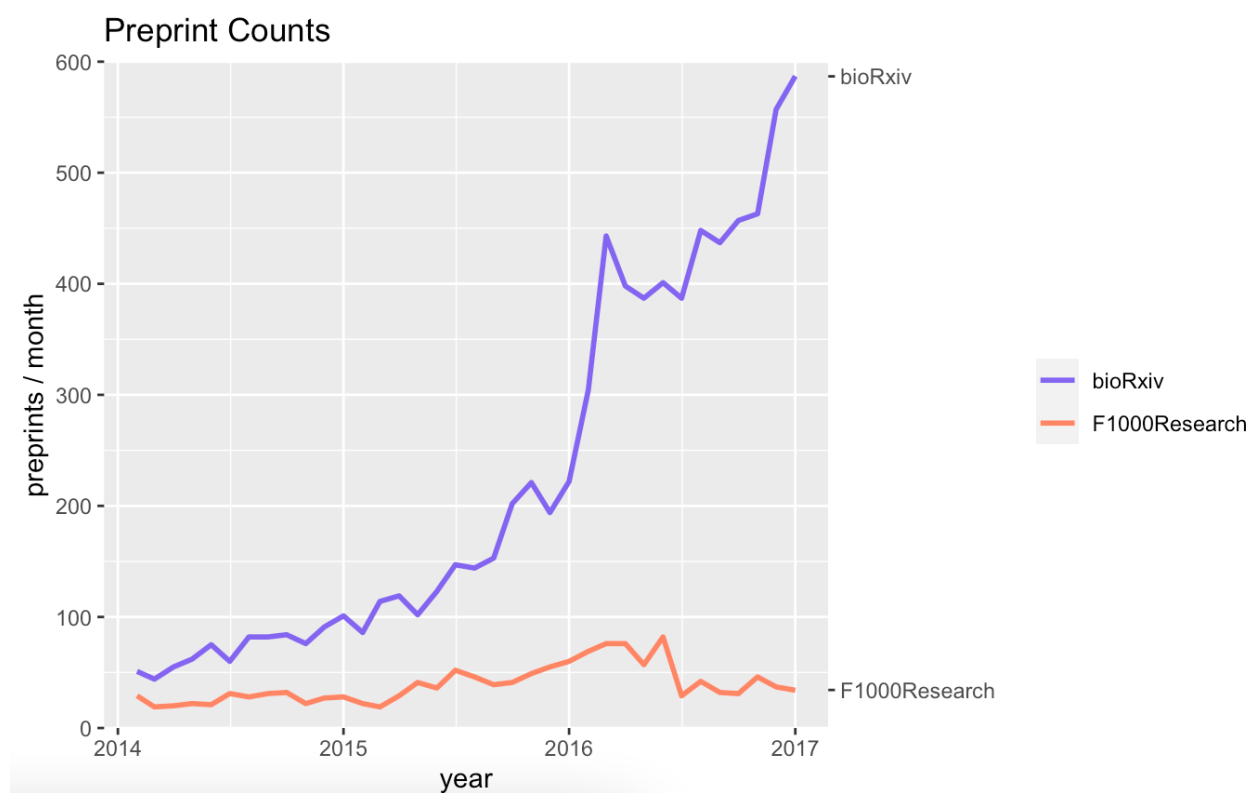
```
1   library(dplyr)
2
3   load("/Volumes/GoogleDrive-102679206261466047710/My Drive/2021-2022/Winter 2022/Data Visualization CSC302/Rscripts/Slides12Rscripts_Data
4   #please change the path if needed
5   head(preprint_growth)
6
7
8
9   preprint_growth %>% filter(archive == "bioRxiv") %>%
10    filter(count > 0) %>% drop_na() -> biorxiv_growth
11  preprint_full<-preprint_growth %>% filter(archive %in%
12                                  c("bioRxiv", "F1000Research")) %>%filter(count > 0) %>%
13    mutate(archive = factor(archive, levels = c("bioRxiv", "F1000Research"))) %>% drop_na() %>% filter(date >= "2004-01-01")
14  preprints_final <- filter(preprints, date == ymd("2017-01-01"))
15
16  ggplot(preprint_full) +
17    aes(date, count, color = archive, fill = archive) +
18    geom_line(size = 1) +
19    scale_y_continuous(
20      limits = c(0, 600), expand = c(0, 0),
21      name = "preprints / month",
22      sec.axis = dup_axis( #this part is for the second y axis
23        breaks = preprints_final$count, #and we use the counts to position our labels
24        labels = c("F1000Research", "bioRxiv"),
25        name = NULL)
26    )+
27    scale_x_date(name = "year",
28                 limits = c(min(ymd("2014-02-01")), ymd("2017-01-01"))) +
29    scale_color_manual(values = c("#7c6bea", "#fe8d6d"),
30                       name = NULL) +
31    theme(legend.position = "right")+
32    ggtitle("Preprint Counts")
33
34
```