Homework 8

Chris Gahagan

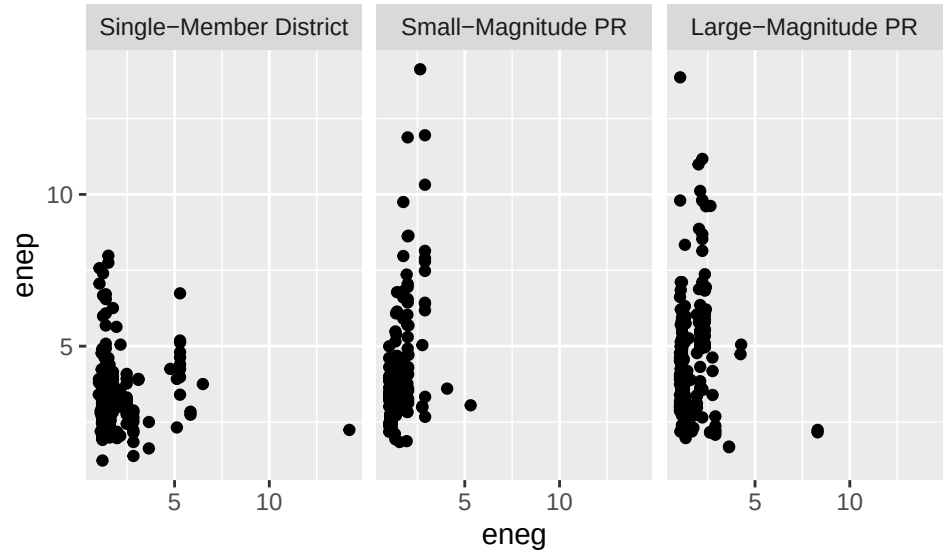
October 21, 2020

Chapter 8 Scatter Plots

Exercise 8.1

Clark and Golder Script

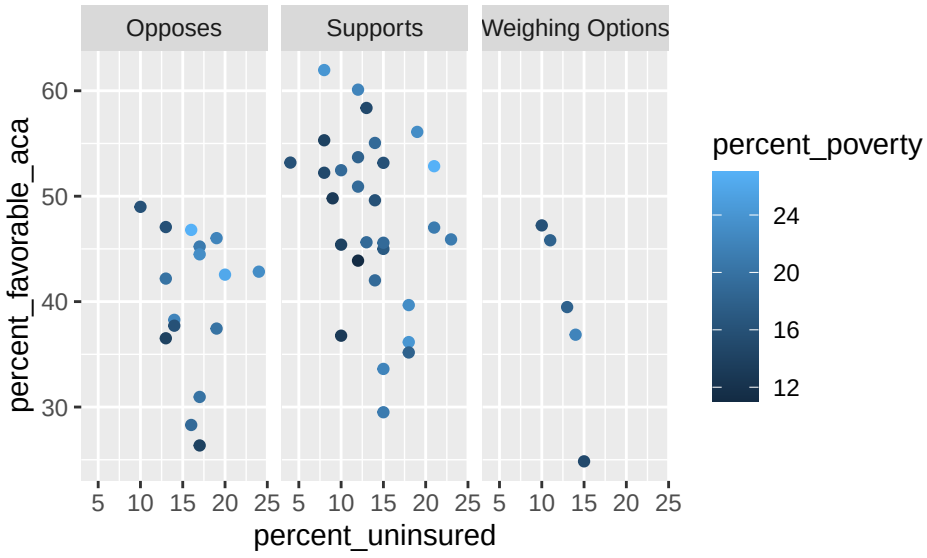
```
parties<- read_rds("data/parties.rds")
ggplot(parties, aes(x = eneg, y = enep)) +
  geom_point()+
  facet_wrap(vars(electoral_system))
```



Exercise 8.2

Uninsured voters ACA favorability by Governor's Position and Poverty Level

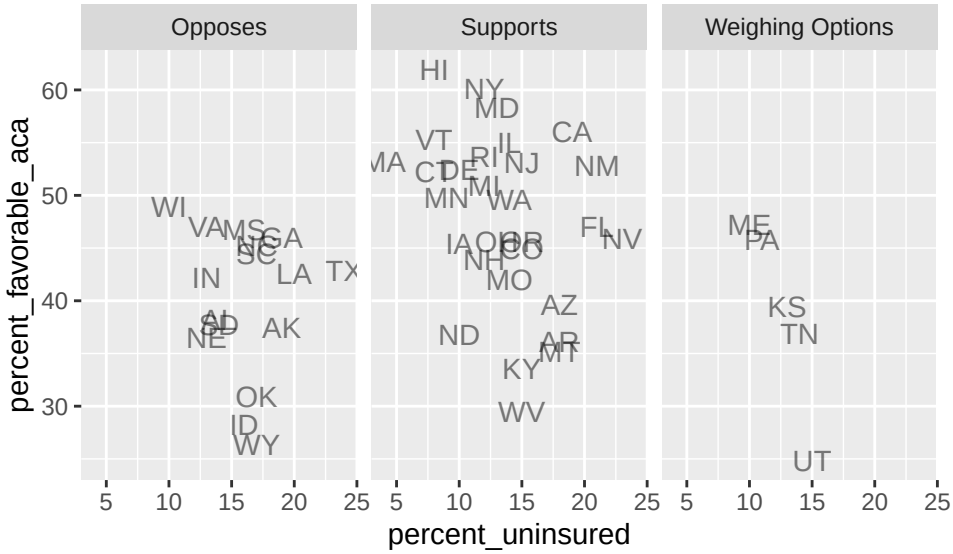
```
politics<-read.csv("data/politics_and_need.csv")
ggplot(politics, aes(x = percent_uninsured, y = percent_favorable_aca, color = percent_poverty)) +
  geom_point()+
facet_wrap(vars(gov_position))
```



Exercise 8.3

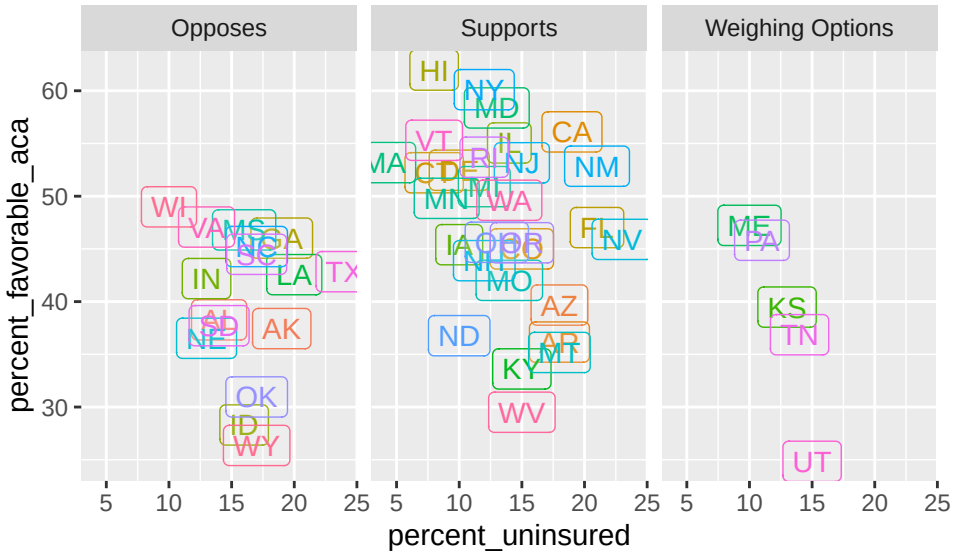
8.3.1 Geom_text() plot

```
ggplot(politics, aes(x = percent_uninsured, y = percent_favorable_aca, label = state_abbr)) +
  geom_text(alpha = .5)+
  facet_wrap(vars(gov_position))
```



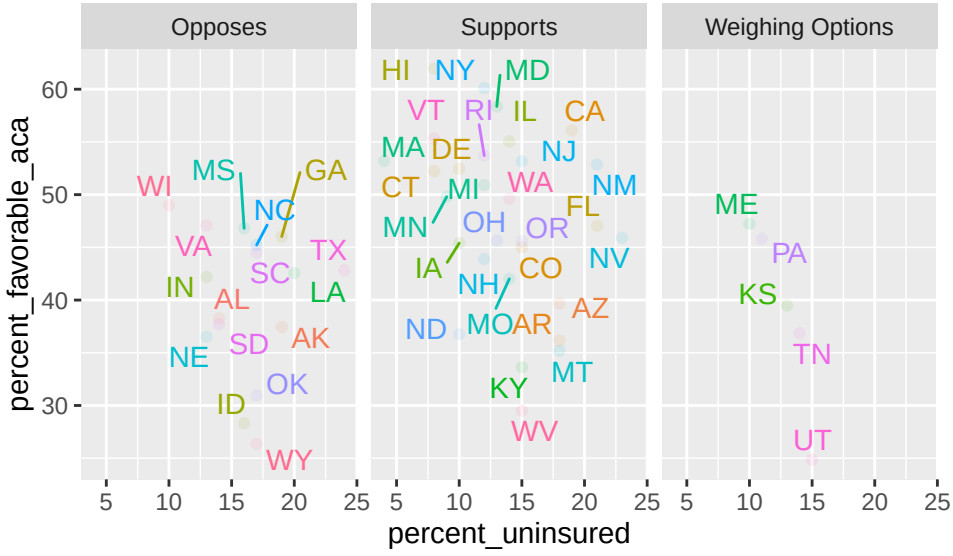
8.3.2 Geom_label() plot

```
ggplot(politics, aes(x = percent_uninsured, y = percent_favorable_aca, label = state_abbr,
                     color = state)) +
  theme(legend.position = "none")+
  geom_label(alpha = .1)+
  facet_wrap(vars(gov_position))
```



8.3.3 Geom_text_repel() plot

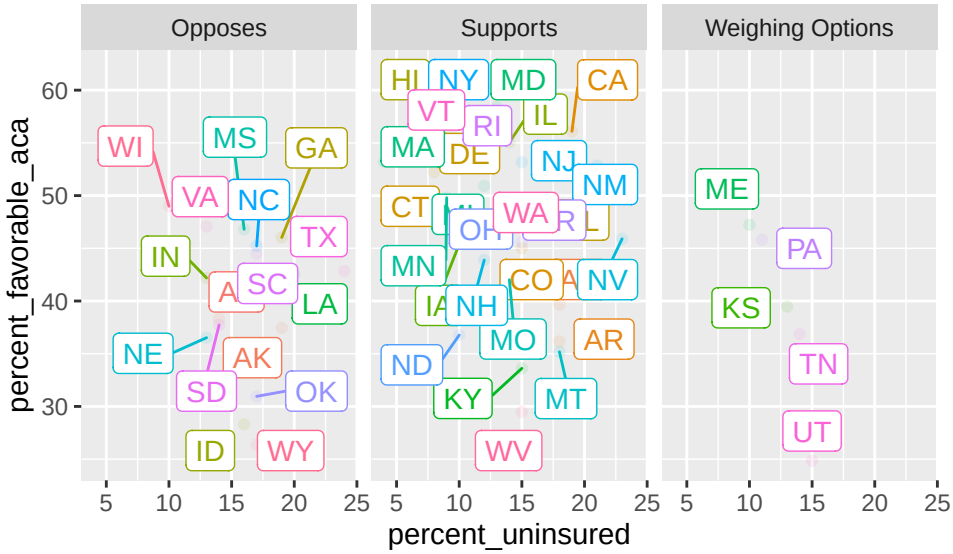
```
ggplot(politics, aes(x = percent_uninsured, y = percent_favorable_aca, label = state_abbr,
                     color = state)) +
  theme(legend.position = "none")+
  geom_point(alpha = .1)+
  geom_text_repel()+
  facet_wrap(vars(gov_position))
```



8.3.4 Geom_label_repel() plot

```
ggplot(politics, aes(x = percent_uninsured, y = percent_favorable_aca, label = state_abbr,
                     color = state)) +
  theme(legend.position = "none")+
  geom_point(alpha = .1)+
```

```
  geom_label_repel()+
  facet_wrap(vars(gov_position))
```



Chapter 9 Correlation Coefficient

Exercise 9.4 Gamson Correlation Data and Plot

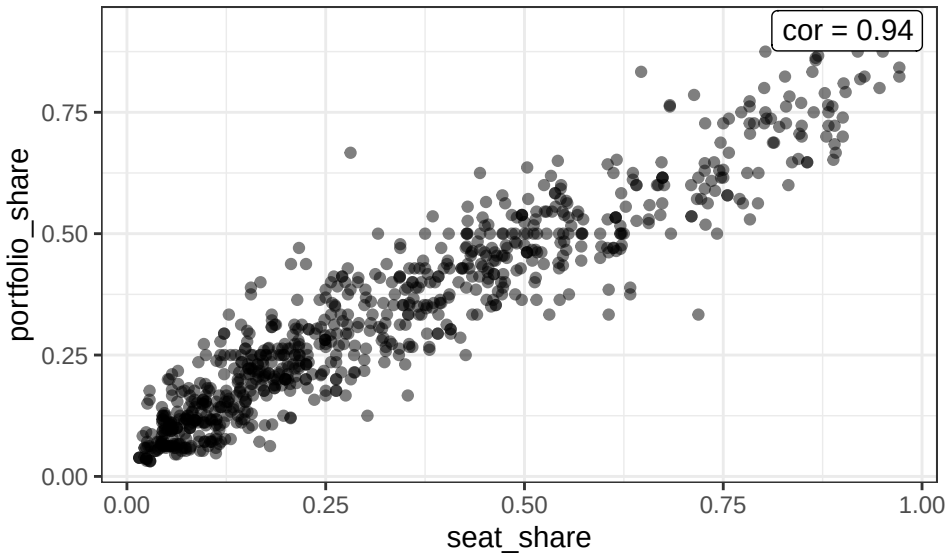
```
gamson<-read_rds("data/gamson.rds") #loading gamson
glimpse(gamson)
```

```
## Rows: 826
## Columns: 2
## $ seat_share      <dbl> 0.02424242, 0.46060607, 0.51515150, 0.47204968, 0.5...
## $ portfolio_share <dbl> 0.09090909, 0.36363637, 0.54545456, 0.45454547, 0.5...
```

```
cor_df<-gamson%>% #creating correlation data
  summarise(cor = cor(seat_share, portfolio_share))
glimpse(cor_df)
```

```
## Rows: 1
## Columns: 1
## $ cor <dbl> 0.9423176
```

```
ggplot(gamson, aes(x = seat_share, y = portfolio_share)) +
  geom_point(alpha = .5)+
geom_label(data = cor_df, aes(x = Inf, y = Inf, label = paste0("cor = ", round(cor, 2))),
           hjust = 1.1, vjust = 1.1)+
  theme_bw() #plotting correlation data
```



Exercise 9.5

9.5.1

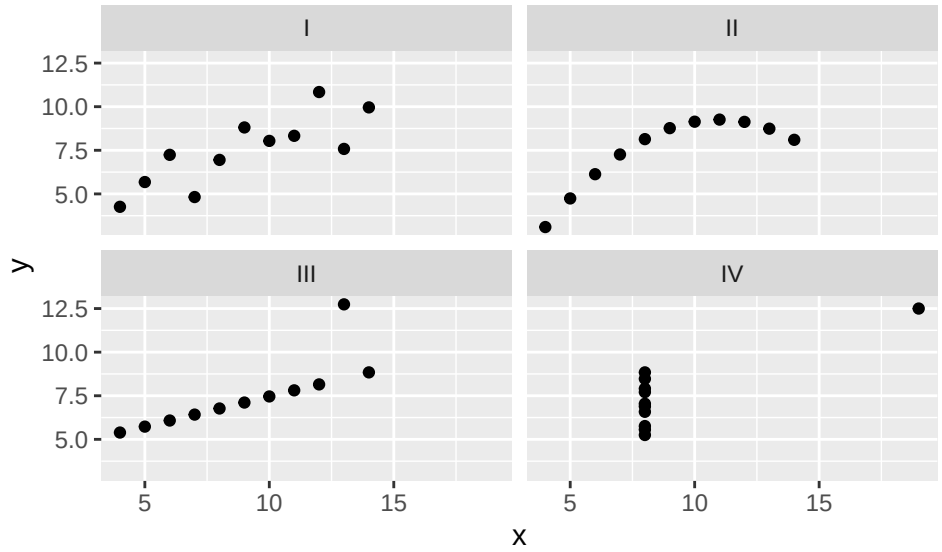
```
devtools::install_github("pos5737/pos5737data")
data(anscombe, package = "pos5737data")

smry_ans <- anscombe %>%
group_by(dataset) %>%
summarize(cor = cor(x = x,
                    y = y))%>%
glimpse()
```

```
## Rows: 4
## Columns: 2
## $ dataset <chr> "I", "II", "III", "IV"
## $ cor     <dbl> 0.8164205, 0.8162365, 0.8162867, 0.8165214
```

1. **All four sets have almost identical correlation**
2. **The data sets each suggest a strong relationship between x and y**

9.5.2 Plots of anscombe

```
ggplot(anscombe, aes(x = x, y = y)) +
  geom_point() +
  facet_wrap(vars(dataset))
```



1. **All the plots are completely different from one another.**
2. **For the amount of data, there seems to be a positive relationship between X and Y, except for in Panel IV, and Panel II the data looks like it is starting to turn negative.**
3. **No, just the correlation coefficient alone did not offer a good summary at all comparably.**