Problem Set 2

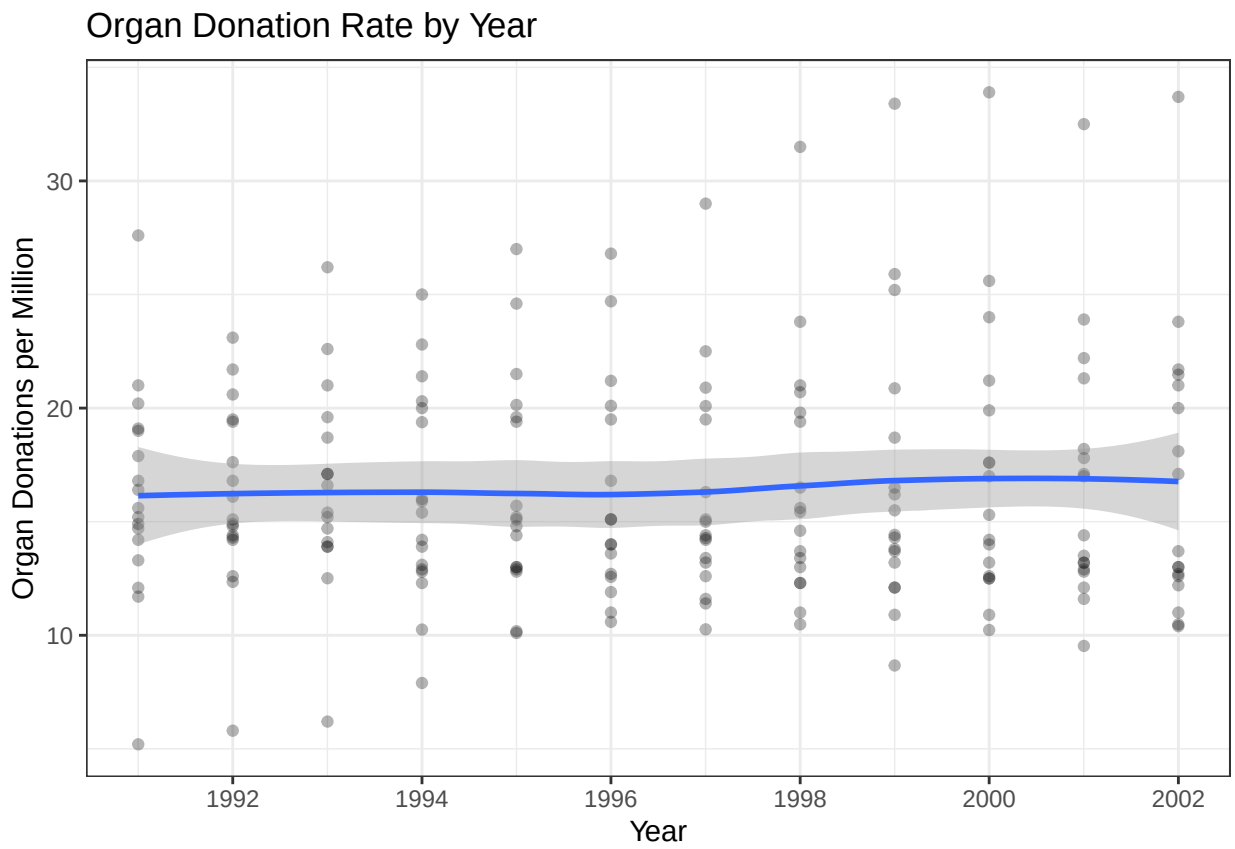
Dav King

2/11/2022

```
library(tidyverse)
library(socviz)
library(cowplot)
library(viridis)
```

Problem 1

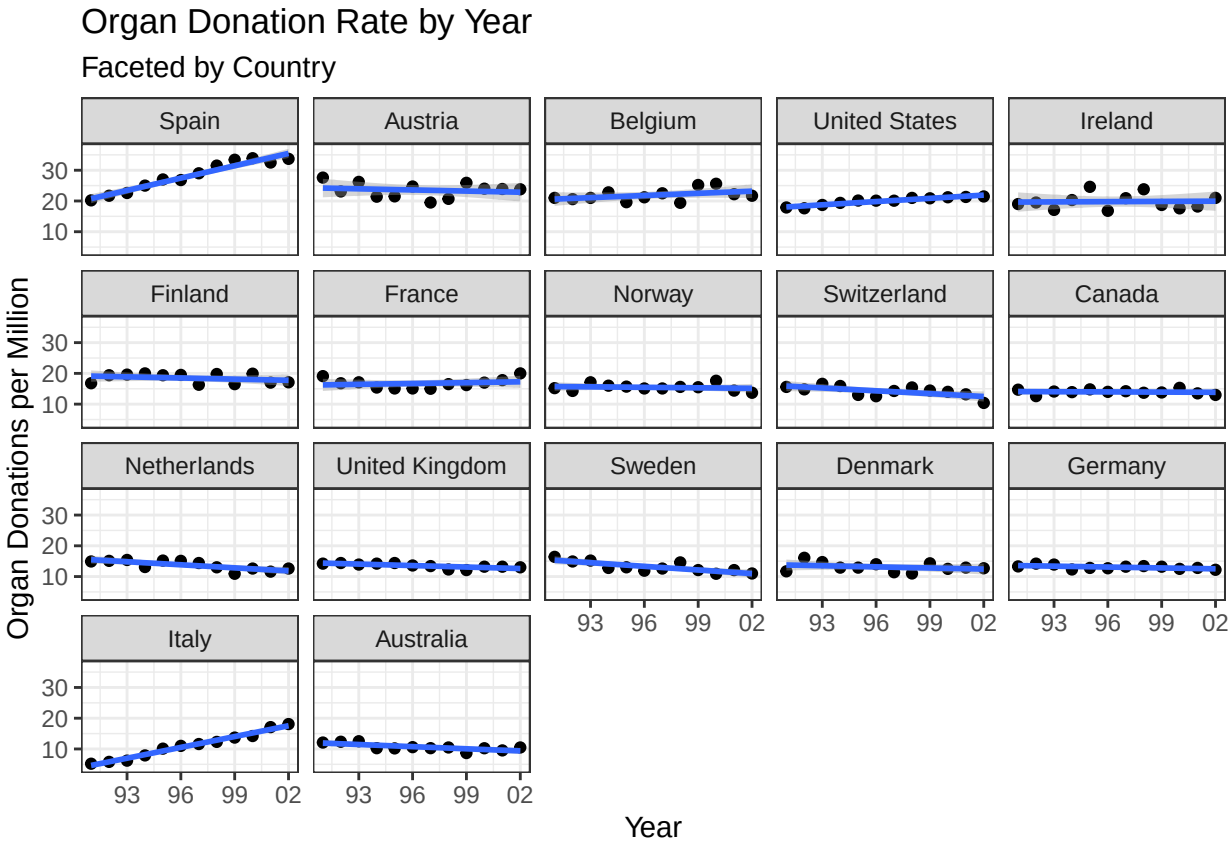
```
ggplot(organdata, aes(x = year, y = donors)) +
  geom_point(alpha = 0.3) +
  geom_smooth() +
  labs(title = "Organ Donation Rate by Year", x = "Year",
       y = "Organ Donations per Million") +
  theme_bw()
```



This is not a super useful way to look at data - I am personally rather opposed to using scatterplots for any data displayed categorically like these years are, and this also reveals no overall trend in the data whatsoever (though there is potentially something of relative interest in the highest donation rate per year, especially if that is only one specific country).

Problem 2

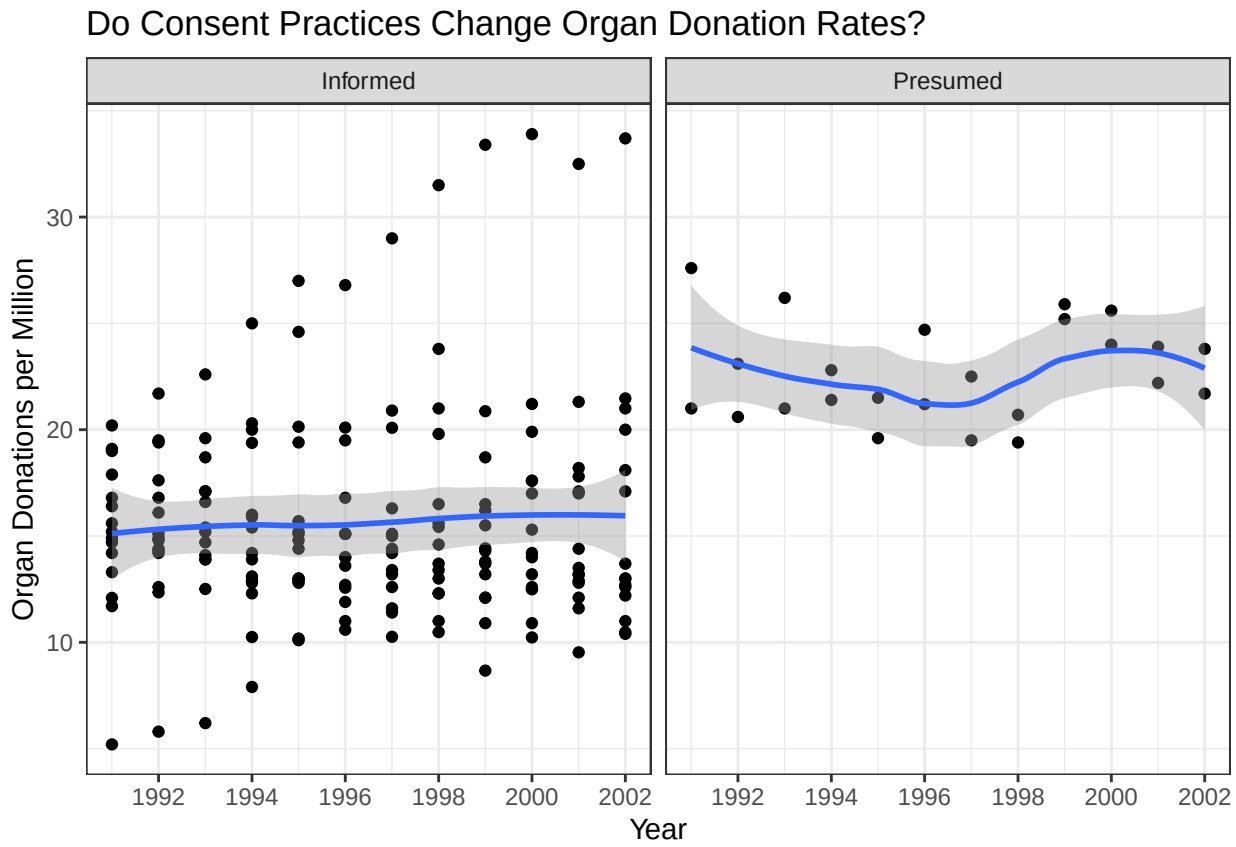
```
ggplot(organdata, aes(x = year, y =donors)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~ reorder(country, -donors, mean, na.rm = T)) +
  labs(title = "Organ Donation Rate by Year", subtitle = "Faceted by Country",
       x = "Year", y = "Organ Donations per Million") +
  theme_bw() +
  scale_x_date(date_breaks = "3 years", date_labels = "%y")
```



I personally do not think using a linear regression line instead of a loess one makes these visualizations any better - in fact, I would argue it makes them worse. Some of these data seem like they might actually have trends that are modeled in nonlinear ways (admittedly, with so few data points it's hard to say whether any regression line is meaningful), and the linear line simply makes things fit just a little more poorly.

Problem 3

```
ggplot(organdata, aes(x = year, y = donors)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(~ consent_practice) +
  labs(title = "Do Consent Practices Change Organ Donation Rates?",
       x = "Year", y = "Organ Donations per Million") +
  theme_bw()
```

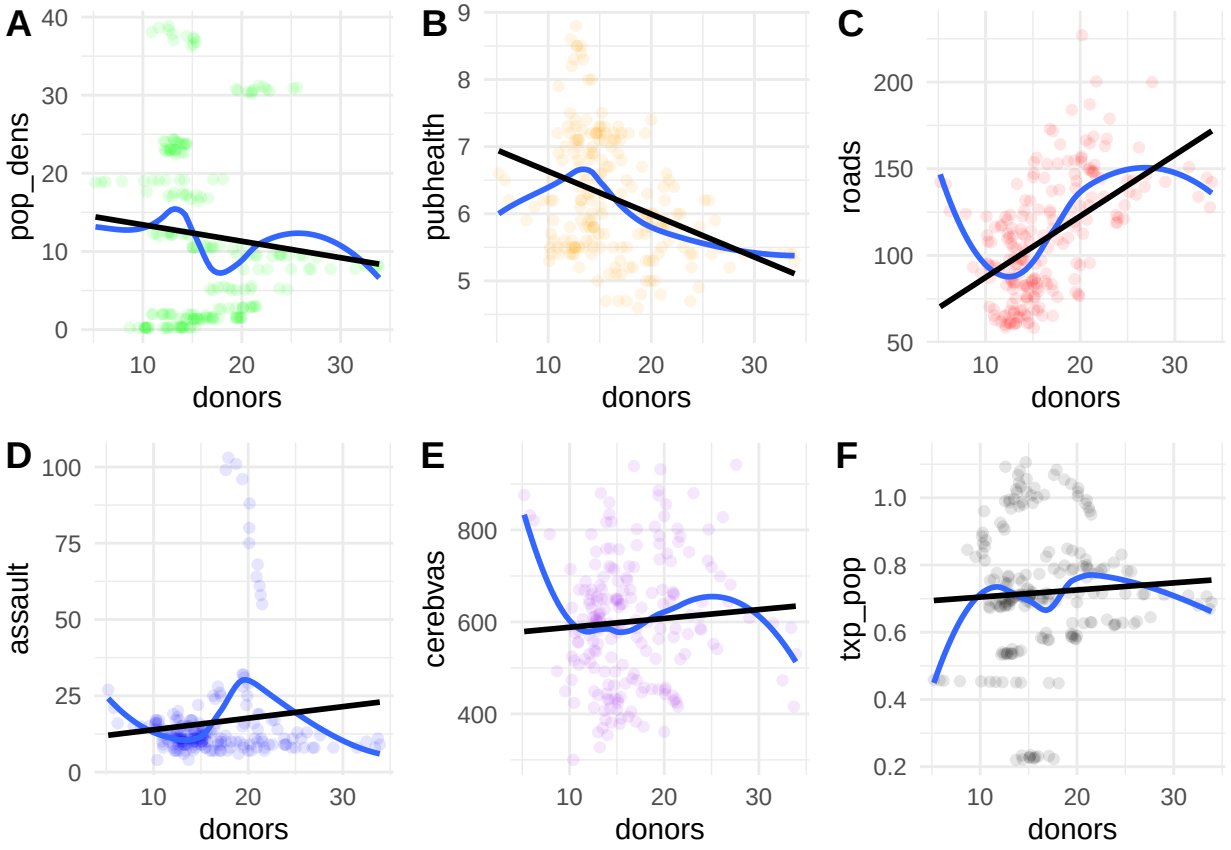


There is a clear difference in the means between these two groups - with countries that presume consent averaging a rate a little over 20/million, and countries that require informed consent averaging closer to 15. The rates do not, however, vary significantly over time in different ways for the two groups (both had relatively constant rates across the sample). I'm still not convinced there isn't an underlying structure in the informed consent countries that I haven't identified yet.

Problem 4

```
p <- ggplot(organdata, aes(x = donors))

a <- p + geom_point(aes(y = pop_dens), alpha = 0.1, color = "green") +
  geom_smooth(aes(y = pop_dens), method = "loess", se = F) +
  geom_smooth(aes(y = pop_dens), method = "lm", color = "black", se = F) +
  theme_minimal()

b <- p + geom_point(aes(y = pubhealth), alpha = 0.1, color = "orange") +
  geom_smooth(aes(y = pubhealth), method = "loess", se = F) +
  geom_smooth(aes(y = pubhealth), method = "lm", color = "black", se = F) +
  theme_minimal()

c <- p + geom_point(aes(y = roads), alpha = 0.1, color = "red") +
  geom_smooth(aes(y = roads), method = "loess", se = F) +
  geom_smooth(aes(y = roads), method = "lm", color = "black", se = F) +
  theme_minimal()
```

```
d <- p + geom_point(aes(y = assault), alpha = 0.1, color = "blue") +
  geom_smooth(aes(y = assault), method = "loess", se = F) +
  geom_smooth(aes(y = assault), method = "lm", color = "black", se = F) +
  theme_minimal()

e <- p + geom_point(aes(y = cerebvas), alpha = 0.1, color = "purple") +
  geom_smooth(aes(y = cerebvas), method = "loess", se = F) +
  geom_smooth(aes(y = cerebvas), method = "lm", color = "black", se = F) +
  theme_minimal()

f <- p + geom_point(aes(y = txp_pop), alpha = 0.1) +
  geom_smooth(aes(y = txp_pop), method = "loess", se = F) +
  geom_smooth(aes(y = txp_pop), method = "lm", color = "black", se = F) +
  theme_minimal()

plot_grid(a, b, c, d, e, f,
          labels = c("A","B", "C", "D", "E", "F"), ncol = 3, nrow = 2)
```



These data continue to hint at underlying structures which are not being revealed properly within this plot. There are lots of clusters of data concentrated around certain points, likely in part because these are not segregated by year. Public health spending is actually associated more with lower donation rates, while transplant donor population is similarly not highly associated with donation rate. Donation rates do also tend to be associated with lower assault cases, however, suggesting that there may be some trend in which countries that have more money and are more devoted to public health are for some reason lower in terms of organ donation rates.

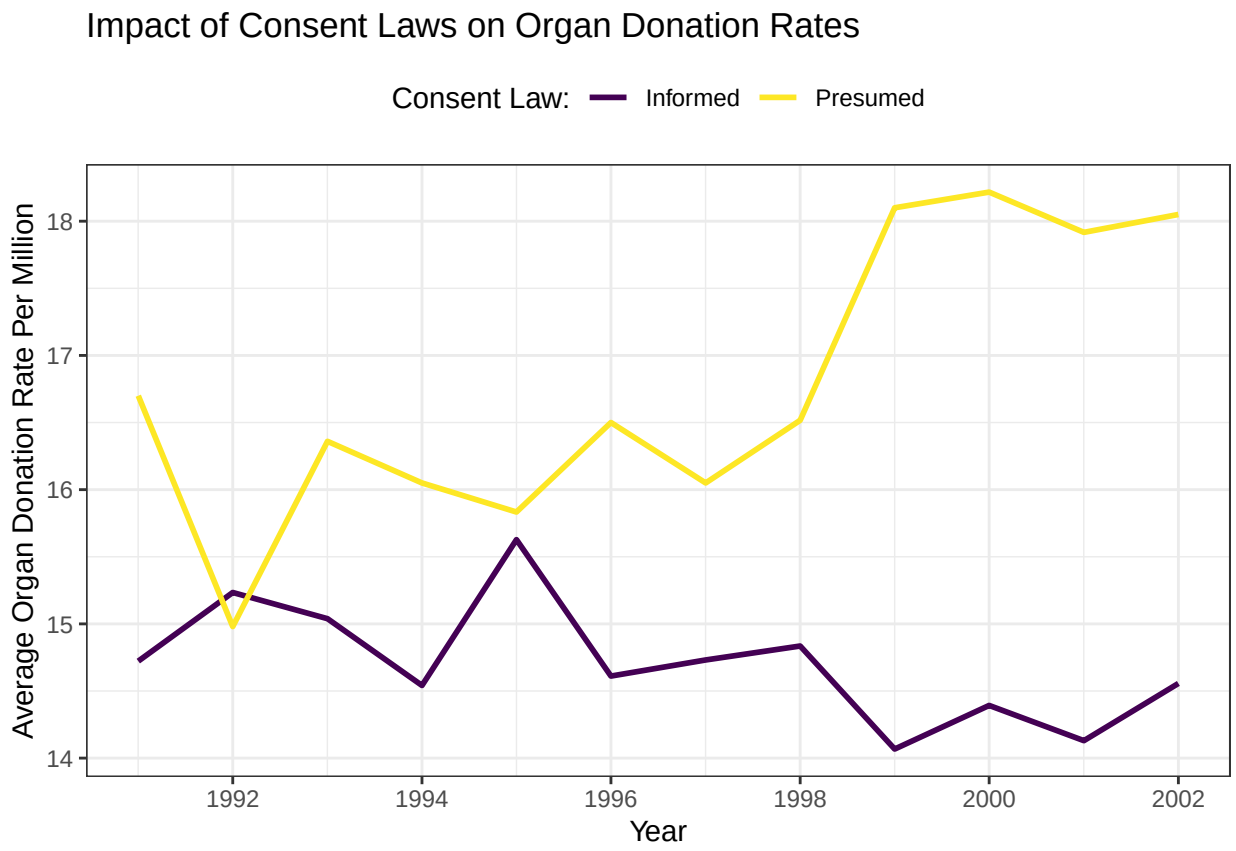
Problem 5

```
organdata %>%
  na.omit(donors) %>%
  group_by(consent_law) %>%
  summarize(avg_don_rate = mean(donors))
```

```
## # A tibble: 2 x 2
##   consent_law avg_don_rate
##   <chr>              <dbl>
## 1 Informed            14.7
## 2 Presumed            16.8
```

Problem 6

```
organdata %>%
  na.omit(donors) %>%
  group_by(consent_law, year) %>%
  summarize(avg_don_rate = mean(donors)) %>%
  ggplot(aes(x = year, y = avg_don_rate, color = consent_law)) +
  geom_line(size = 1) +
  scale_color_viridis(discrete = T) +
  labs(title = "Impact of Consent Laws on Organ Donation Rates", x = "Year",
       y = "Average Organ Donation Rate Per Million", color = "Consent Law:") +
  theme_bw() +
  theme(legend.position = "top")
```



Problem 7

```
organdata %>%
  na.omit(donors) %>%
  group_by(consent_law, year) %>%
  summarize(avg_don_rate = mean(donors), sd_don_rate = sd(donors)) %>%
  ggplot(aes(x = year, y = avg_don_rate)) +
  geom_ribbon(aes(ymin = avg_don_rate - sd_don_rate,
                  ymax = avg_don_rate + sd_don_rate,
                  fill = consent_law), alpha = 0.2) +
  geom_line(aes(color = consent_law), size = 1) +
  labs(title = "Impact of Consent Laws on Organ Donation Rates",
       subtitle = "Filled Area = Standard Deviation", x = "Year",
       y = "Average Organ Donation Rate Per Million", color = "Consent Law:") +
  theme_bw() +
  theme(legend.position = "top") +
  guides(fill = "none") +
  scale_color_viridis(discrete = T) +
  scale_fill_viridis(discrete = T)
```

Impact of Consent Laws on Organ Donation Rates

Filled Area = Standard Deviation