portfolio_4

Matthew Voss

5/2/2022

source: https://www.kaggle.com/yamqwe/pregnancy-birth-abortion-rates-in-the-united-stae

```
####Here is my original code, the new stuff is below ########
df=read_csv("State Pregnancy-Birth-Abortion Rates.csv")
```

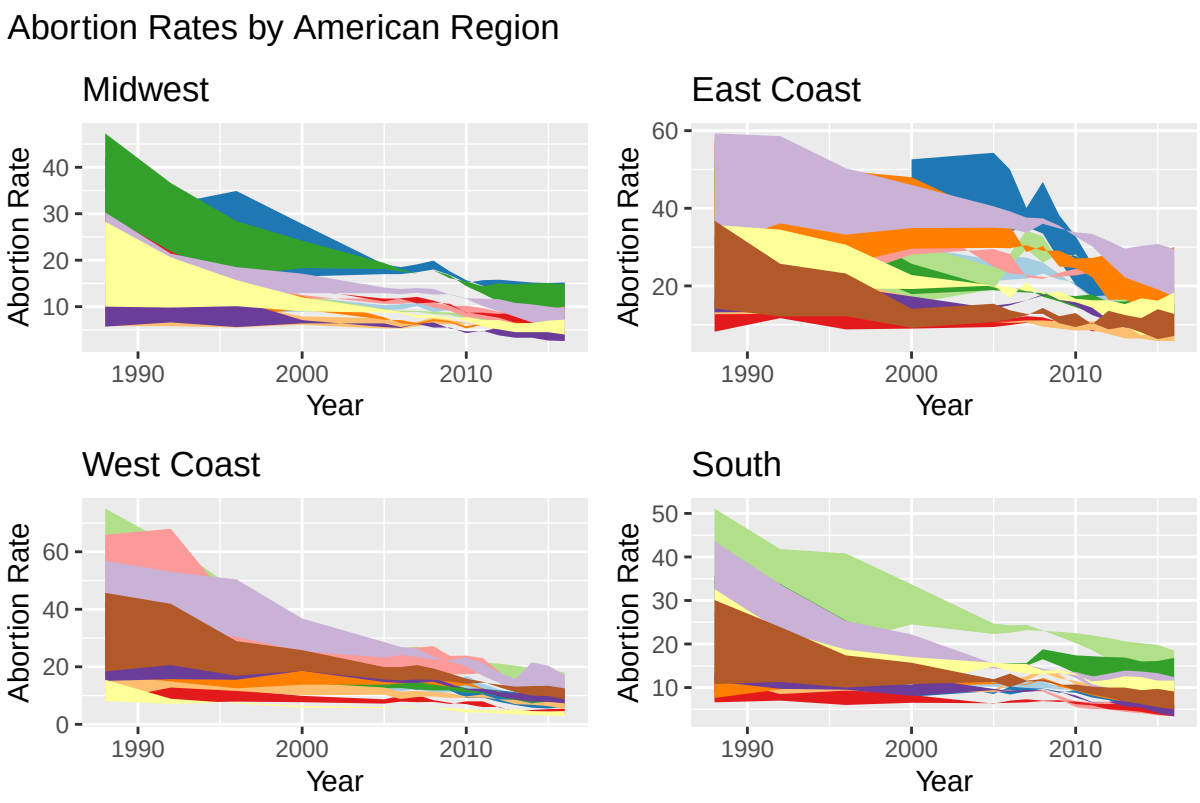
```
## Rows: 23172 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): State, Metric, Age Range
## dbl (2): Year, Events per 1,000 women
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
south=c("TX","LA","AR","TN","NC","SC","GA","FL","AL","MS","WV","VA","OK","KS","KY")
east=c("ME","MA","VT","NH","RI","CT","NY","NJ","DE","MD","PA","DC")
west=c("WA","OR","CA","ID","MT","WY","CO","UT","NV","AZ","AK","HI","NM")
midwest=c("OH","MI","WI","MN","IA","NE","ND","SD","MO","IL","IN")
sstates=df[df$State %in% south,] %>%
  filter(Metric %in% "Abortion Rate") %>%
  filter(`Age Range`%in%c("15-19","Total","30-34")) %>%
  pivot_wider(names_from = `Age Range`,values_from=`Events per 1,000 women`)
estates=df[df$State%in%east,]%>%
  filter(Metric %in% "Abortion Rate") %>%
  filter(`Age Range`%in%c("15-19","Total","30-34")) %>%
  pivot_wider(names_from = `Age Range`,values_from=`Events per 1,000 women`)
wstates=df[df$State%in%west,]%>%
  filter(Metric %in% "Abortion Rate") %>%
  filter(`Age Range`%in%c("15-19","Total","30-34")) %>%
  pivot_wider(names_from = `Age Range`,values_from=`Events per 1,000 women`)
mstates=df[df$State%in%midwest,]%>%
  filter(Metric %in% "Abortion Rate") %>%
  filter(`Age Range`%in%c("15-19","Total","30-34")) %>%
  pivot_wider(names_from = `Age Range`,values_from=`Events per 1,000 women`)
p=list()
p[["so"]]=ggplot(sstates,aes(x=Year))+geom_ribbon(aes(ymax=`15-19`,ymin=`30-34`,group=State,fill=State))
    scale_fill_brewer(palette="Paired")
p[["we"]]=ggplot(wstates,aes(x=Year))+geom_ribbon(aes(ymax=`15-19`,ymin=`30-34`,group=State,fill=State))
    scale_fill_brewer(palette="Paired")
p[["ea"]]=ggplot(estates,aes(x=Year))+geom_ribbon(aes(ymax=`15-19`,ymin=`30-34`,group=State,fill=State))
    scale_fill_brewer(palette="Paired")
```

```
p[["mw"]]=ggplot(mstates,aes(x=Year))+geom_ribbon(aes(ymax=`15-19`,ymin=`30-34`,group=State,fill=State))
    scale_fill_brewer(palette="Paired")

(p[["mw"]]+p[["ea"]])/(p[["we"]]+p[["so"]])+plot_annotation(title="Abortion Rates by American Region",ca
```

```
## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Paired is 12
## Returning the palette you asked for with that many colors
```

```
## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Paired is 12
## Returning the palette you asked for with that many colors
```

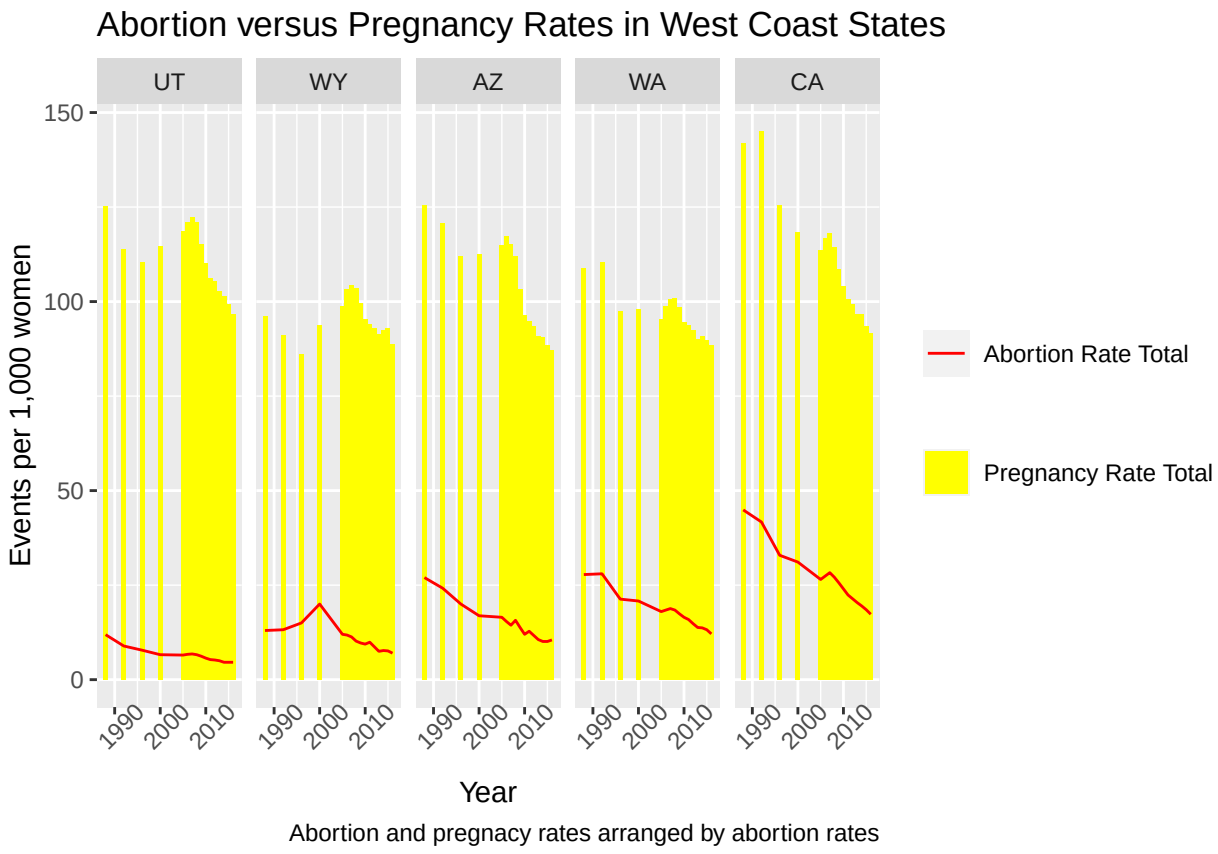


Abortion Rates between the age groups of 15−19 and 30−34 from 1988 to 2016

```
#fig 2:
westCoastAbortionRatesOrder=df %>%
  filter(State%in%west) %>%
  filter(`Age Range`=="Total") %>%
  pivot_wider(names_from = "Metric", values_from = `Events per 1,000 women`) %>%
  group_by(State) %>%
  summarise(mu=mean(`Abortion Rate`)) %>%
  arrange(mu)
#Quantiles of Western States: UT,Q1: WY, Q2: AZ, Q3: WA, CA
second=df %>%
  filter(State%in%c("UT","WY","AZ","WA","CA")) %>%
  filter(`Age Range`=="Total") %>%
  pivot_wider(names_from = "Metric", values_from = `Events per 1,000 women`) %>%
  left_join(westCoastAbortionRatesOrder)
```

```
## Joining, by = "State"
```

```
ggplot(second)+geom_col(aes(Year,`Pregnancy Rate`,fill="Pregnancy Rate Total"))+geom_line(aes(x=Year,y=
  facet_grid(~reorder(State,mu))+
  scale_fill_manual(name=NULL,values=c("Pregnancy Rate Total"="Yellow"))+
  scale_color_manual(name=NULL,values = c("Abortion Rate Total"="red"))+
  labs(title="Abortion versus Pregnancy Rates in West Coast States",y="Events per 1,000 women",caption =
  theme(axis.text.x = element_text(angle=45))
```



Abortion and pregnacy rates arranged by abortion rates

```
p<-list()
temp_data<-df %>%
  filter(Metric%in%c("Abortion Rate","Pregnancy Rate")) %>%
  mutate(region = ifelse(State%in%south,"South",ifelse(State%in%east,"East Coast",ifelse(
  State%in%west,"West Coast","Midwest")))) %>%
  filter(`Age Range`=="Total") %>%
  filter(State!="US") %>%
  mutate(State = ifelse(State=="DC","District of Colombia",State)) %>%
  pivot_wider(names_from = Metric, values_from = `Events per 1,000 women`) %>%
  rename(ID=State) %>%
  mutate(ID=abbr2state(ID)) %>%
  drop_na()
bidata<-bi_class(temp_data,x=`Abortion Rate`,y=`Pregnancy Rate`,style="jenks",dim=3)
plot<-states %>%
    left_join(bidata) %>%
    drop_na() %>%
    ggplot()+
```

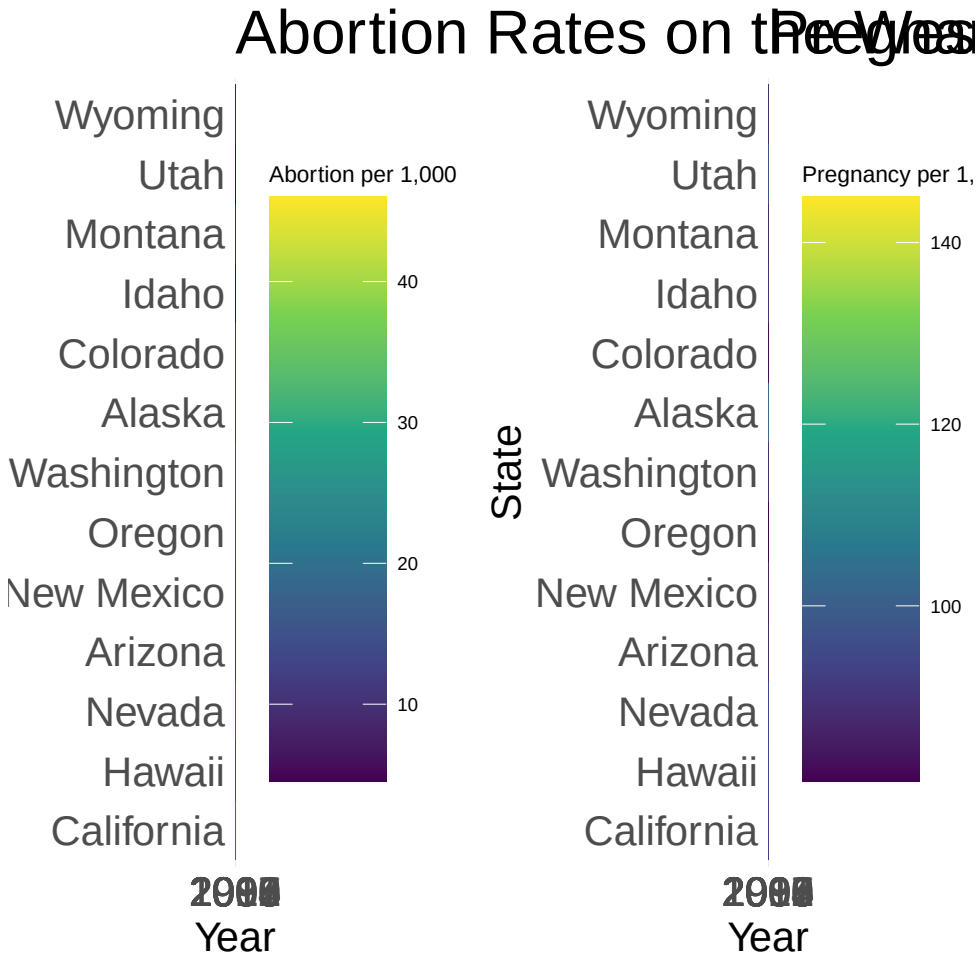
```
    geom_sf(mapping = aes(fill=bi_class),show.legend = F)+
    bi_scale_fill(pal = "DkViolet")+
    labs(title = "Abortion Rates vs Pregnancy Rates by State", subtitle = "Aggregated from 1988 to 2016
    theme_void()+
    theme(plot.title = element_text(hjust=0.5),
          plot.subtitle = element_text(hjust=0.5)) +
    geom_label(aes(X,Y,label=ID),size=3.5)
```

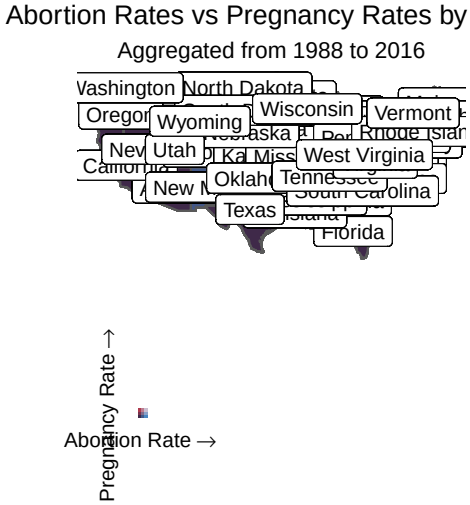
```
## Joining, by = "ID"
```

```
legend<-bi_legend(pal = "DkViolet",dim=3,xlab = "Abortion Rate", ylab = "Pregnancy Rate")
#png("portfolio4_1.png",width=900,height = 900)
p[["bi"]]<-ggdraw()+
    draw_plot(plot,0.02)+
    draw_plot(legend,0,0,0.25,0.25)
#dev.off()
```

```
####Second Plot, but I wanted to changed the plot away from bar graphs as they were too hard to read
abortion_df<-df %>%
  filter(State%in%west) %>%
  filter(`Age Range`=="Total") %>%
  pivot_wider(names_from = "Metric", values_from = `Events per 1,000 women`) %>%
  mutate(State=abbr2state(State)) %>%
  select(Year,State,`Abortion Rate`) %>%
  pivot_wider(names_from = Year,values_from = c(`Abortion Rate`)) %>%
  column_to_rownames(var="State")

clusters<-abortion_df %>%
  kmeans(centers=5)

abortion<-broom::augment(clusters,abortion_df)%>%
  rename(State=".rownames") %>%
  pivot_longer(cols = -c(State,.cluster),names_to = "Year",values_to = "AbortionRate")


p[["ab"]]<-ggplot(abortion)+
  geom_tile(aes(x=factor(Year),y=factor(reorder(State,as.numeric(.cluster))),fill=AbortionRate))+
  scale_fill_viridis_c()+
  theme_minimal()+
  theme(axis.text.x = element_text(size = 20),
        axis.text.y = element_text(size = 20),
        axis.title.x = element_text(size = 20),
        axis.title.y = element_text(size = 20),
        legend.key.size = unit(2,"cm"),
        plot.title = element_text(size = 30))+
  labs(title = "Abortion Rates on the West Coast",x="Year",y="State",
       fill="Abortion per 1,000")

pregnancy<-df %>%
  filter(State%in%west) %>%
  filter(`Age Range`=="Total") %>%
  pivot_wider(names_from = "Metric", values_from = `Events per 1,000 women`) %>%
  mutate(State=abbr2state(State)) %>%
```

```
  select(Year,State,`Pregnancy Rate`) %>%
  pivot_wider(names_from = Year,values_from = `Pregnancy Rate`) %>%
  column_to_rownames(var="State")

pregnancy<-broom::augment(clusters,pregnancy) %>%
  rename("State"=.rownames) %>%
  pivot_longer(cols = -c(State,.cluster),names_to = "Year",values_to = "Pregnancy")

p[["pr"]]<-ggplot(pregnancy)+
  geom_tile(aes(x=factor(Year),y=factor(reorder(State,as.numeric(.cluster))),fill=Pregnancy))+
  theme_minimal()+
  scale_fill_viridis_c()+
  theme(axis.text.x = element_text(size = 20),
        axis.text.y = element_text(size = 20),
        axis.title.x = element_text(size = 20),
        axis.title.y = element_text(size = 20),
        legend.key.size = unit(2,"cm"),
        plot.title = element_text(size = 30))+
  labs(title="Pregnancy Rates on the West Coast",fill="Pregnancy per 1,000",
       x="Year",y="State")

(p[["ab"]]+p[["pr"]])/p[["bi"]]
```


Abortion Rates on the West
Pregnancy Rates on the West
Wyoming
Utah
Montana
Idaho
Colorado
Alaska
Washington
Oregon
New Mexico
Arizona
Nevada
Hawaii
California
State
Abortion per 1,000
Pregnancy per 1,
40
30
20
10
140
120
100
Year
Year


Abortion Rates vs Pregnancy Rates by State
Aggregated from 1988 to 2016
Wyoming
Wisconsin
Vermont
Utah
West Virginia
Texas
Florida
Pregnancy Rate →
Abortion Rate →

```
#ggsave("portfolio4.png")
```