lab1_samir

Samir Datta

September 25, 2017

```
library(ggplot2)
theme_set(theme_bw())
library(car)
publicopinion <- read.csv('C:/Users/Samir/Documents/MIDS/StatsF17/lab 1/public_opinion.csv')

publicopinion$partyfactor <- ifelse(publicopinion$party==1, 'Democrat',
                                    ifelse(publicopinion$party==2, 'Other',
                                           'Republican'))
publicopinion$age <- 2017 - publicopinion$birthyr
publicopinion$genderfactor <- ifelse(publicopinion$gender==1,'Male', 'Female')
publicopinion$racefactor <- ifelse(publicopinion$race_white==1, 'White', 'Non-White')
publicopinion$spfactor <- ifelse(publicopinion$sanders_preference==1, "Yes", "No")
publicopinion_narm <- publicopinion[!is.na(publicopinion$sanders_preference),]
```

```
po_party_agg <- with(publicopinion_narm,
                     aggregate(cbind(100*sanders_preference),
                               list(partyfactor=partyfactor),
                               mean))
po_party_df <- data.frame(sanders_pref_percent=po_party_agg$V1,
                          party=po_party_agg$partyfactor)
po_party_df
```

```
##   sanders_pref_percent      party
## 1             45.27473   Democrat
## 2             65.93886      Other
## 3             64.02878 Republican
```

```
po_gender_agg <- with(publicopinion_narm,
                     aggregate(cbind(100*sanders_preference),
                               list(genderfactor=genderfactor),
                               mean))
po_gender_df <- data.frame(sanders_pref_percent=po_gender_agg$V1,
                          gender=po_gender_agg$genderfactor)
po_gender_df
```

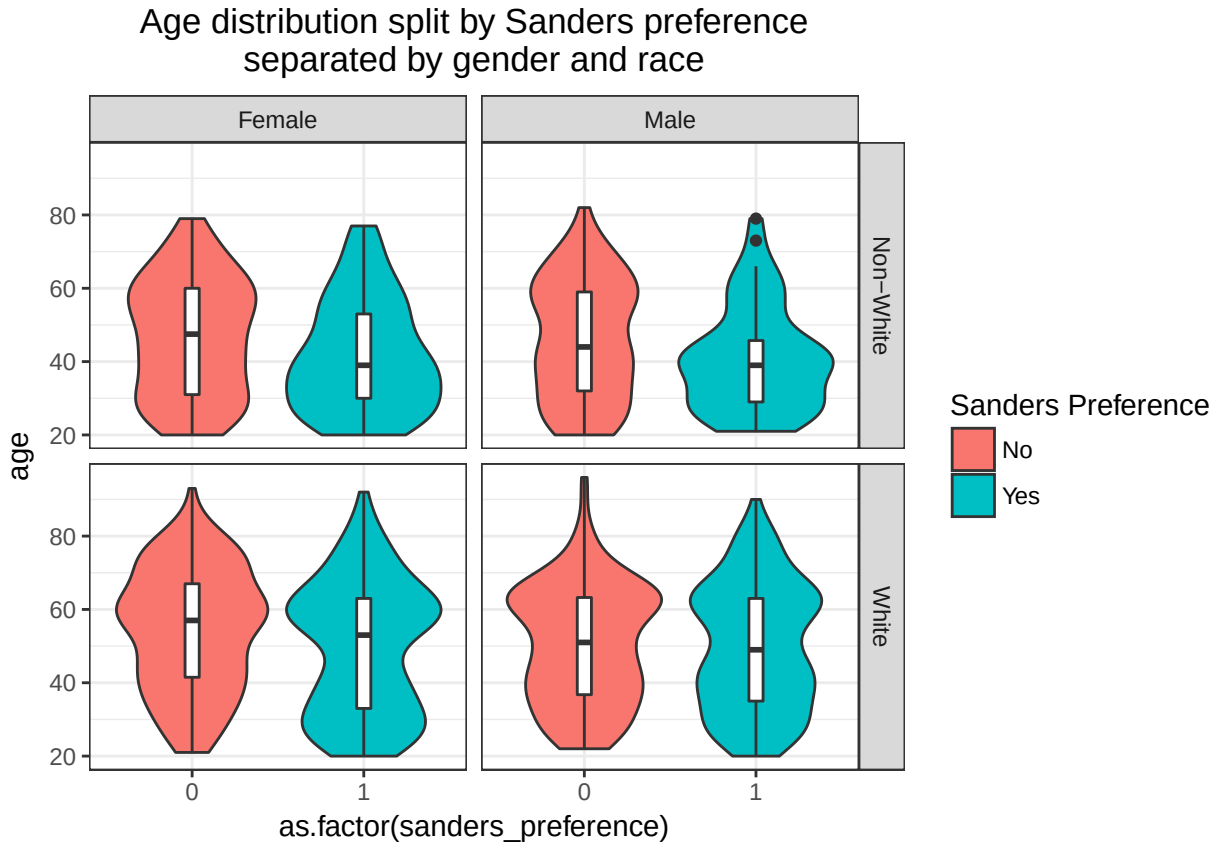
```
##   sanders_pref_percent gender
## 1             57.71704 Female
## 2             57.46924   Male
```

```
po_race_agg <- with(publicopinion_narm,
                     aggregate(cbind(100*sanders_preference),
                               list(racefactor=racefactor),
                               mean))
po_race_df <- data.frame(sanders_pref_percent=po_race_agg$V1,
                          gender=po_race_agg$racefactor)
po_race_df
```

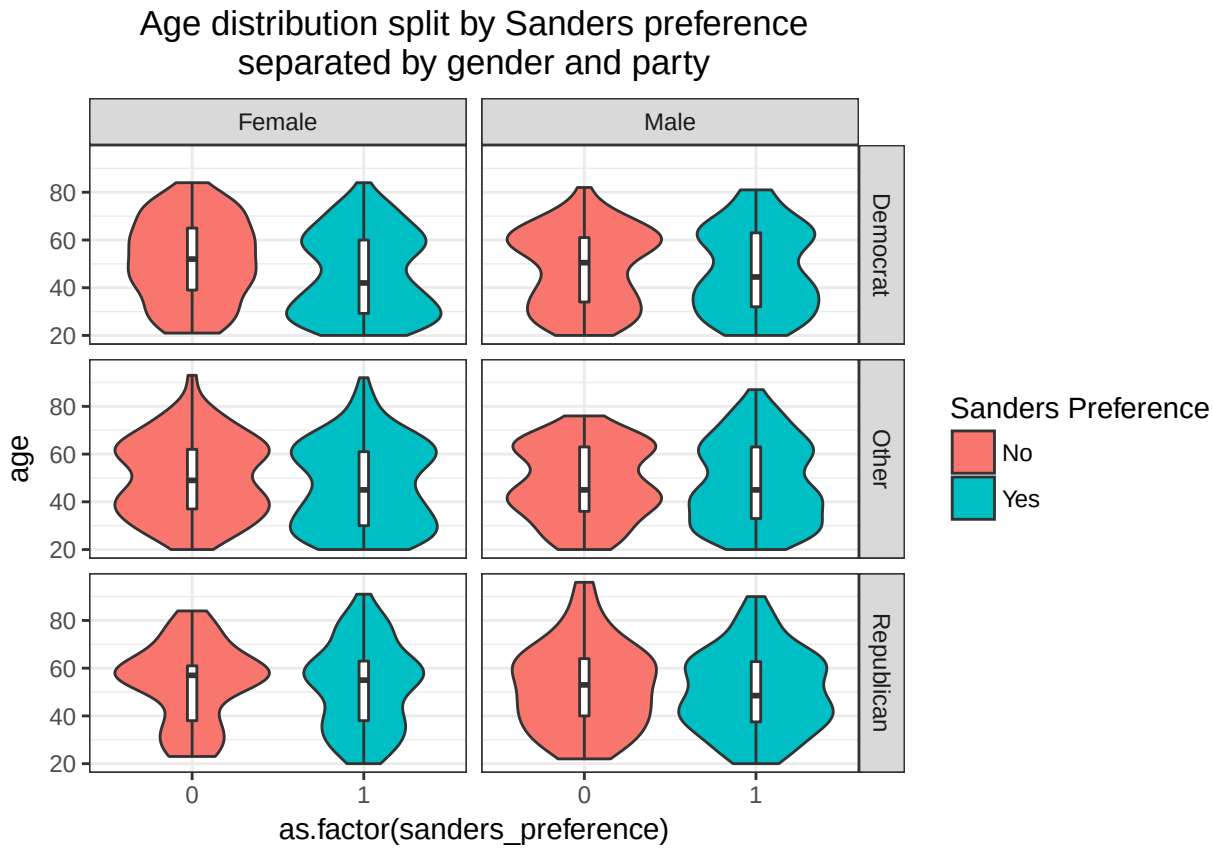
```
##   sanders_pref_percent    gender
## 1             40.99379 Non-White
```

```
## 2             63.75144     White
```

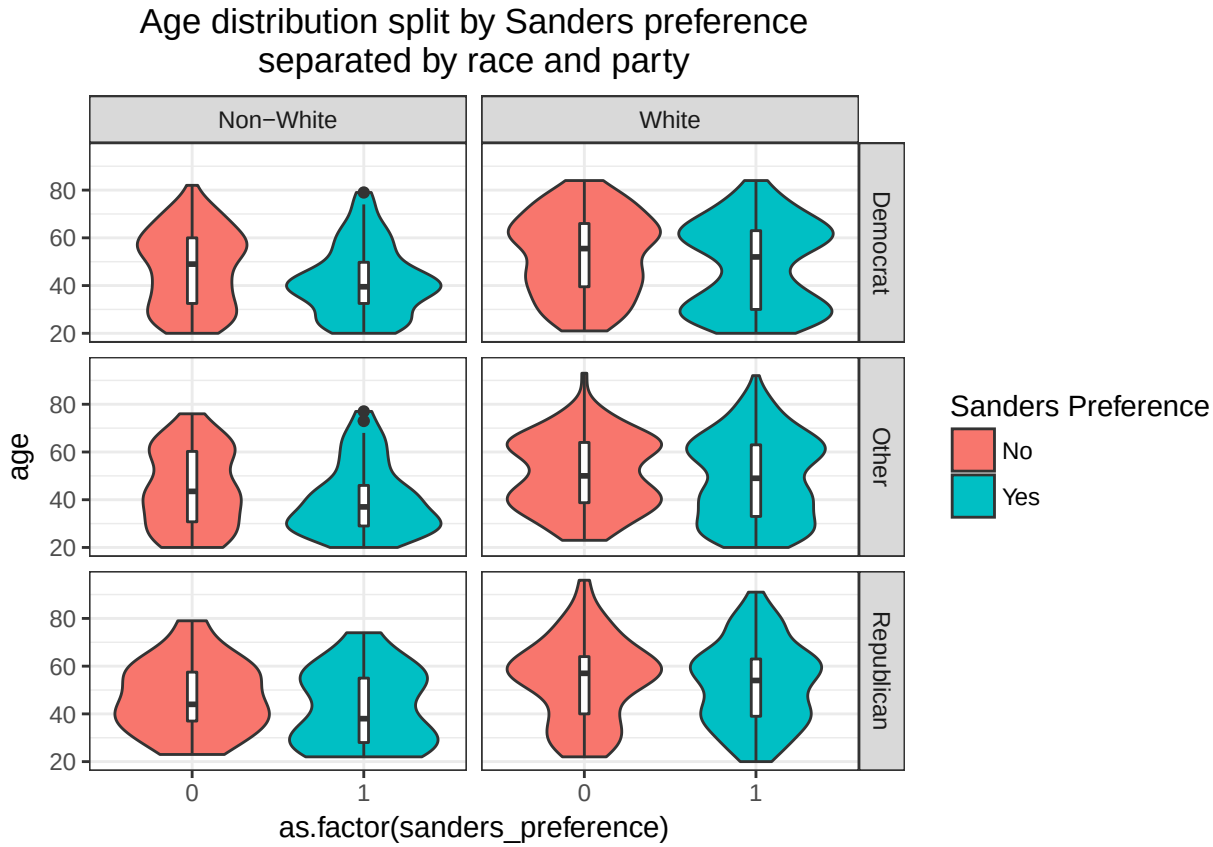
```
ggp <- ggplot(aes(x=as.factor(sanders_preference), y=age,
                  fill=as.factor(sanders_preference)),
              data=publicopinion_narm)

ggp+geom_violin()+geom_boxplot(width=0.33, fill="white")+
  facet_grid(racefactor~genderfactor)+
  scale_fill_discrete(name="Sanders Preference",
                      labels=c("No", "Yes"))+
  ggtitle("Age distribution split by Sanders preference\nseparated by gender and race")+
  theme(plot.title = element_text(hjust = 0.5))
```



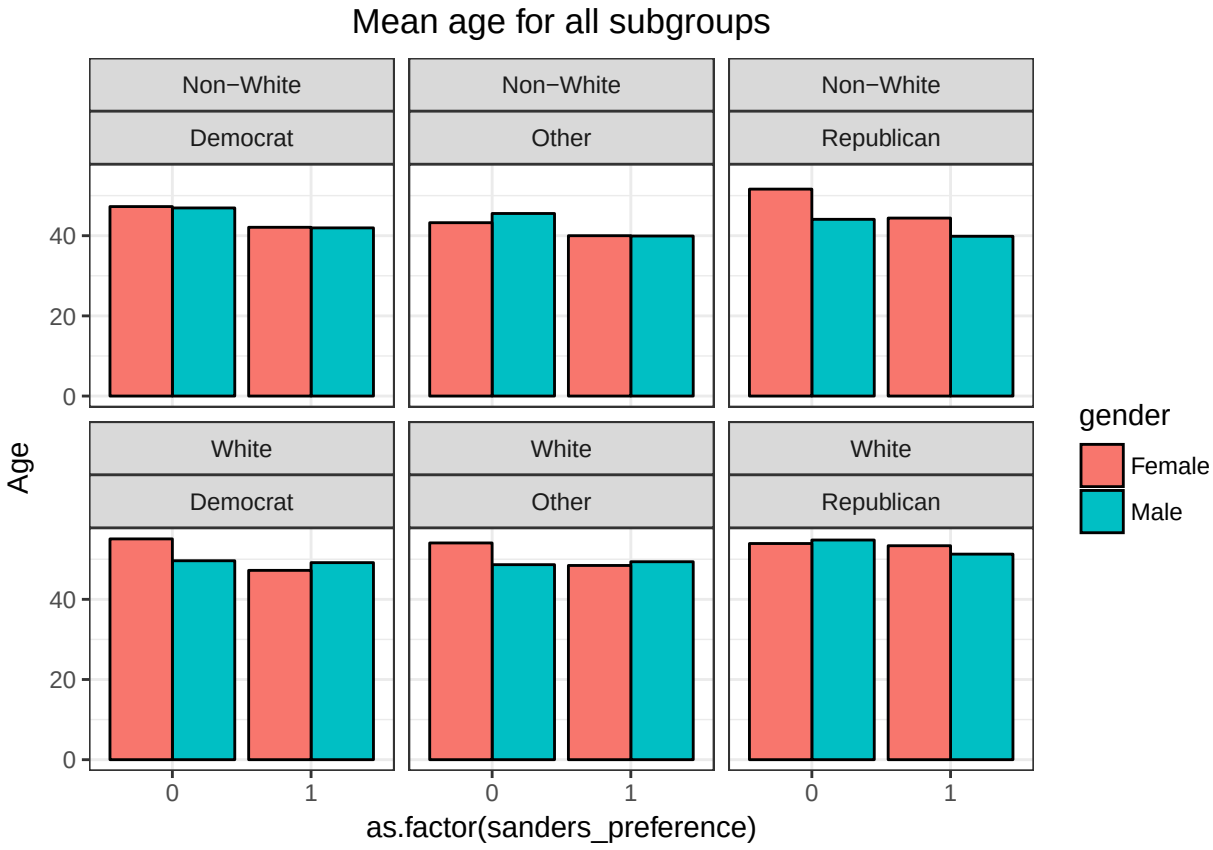

```
ggp <- ggplot(aes(x=as.factor(sanders_preference), y=age,
                  fill=as.factor(sanders_preference)),
              data=publicopinion_narm)

ggp+geom_violin()+geom_boxplot(width=0.33, fill="white")+
  facet_grid(partyfactor~genderfactor)+
  scale_fill_discrete(name="Sanders Preference",
                      labels=c("No", "Yes"))+
  ggtitle("Age distribution split by Sanders preference\nseparated by gender and party")+
  theme(plot.title = element_text(hjust = 0.5))
```



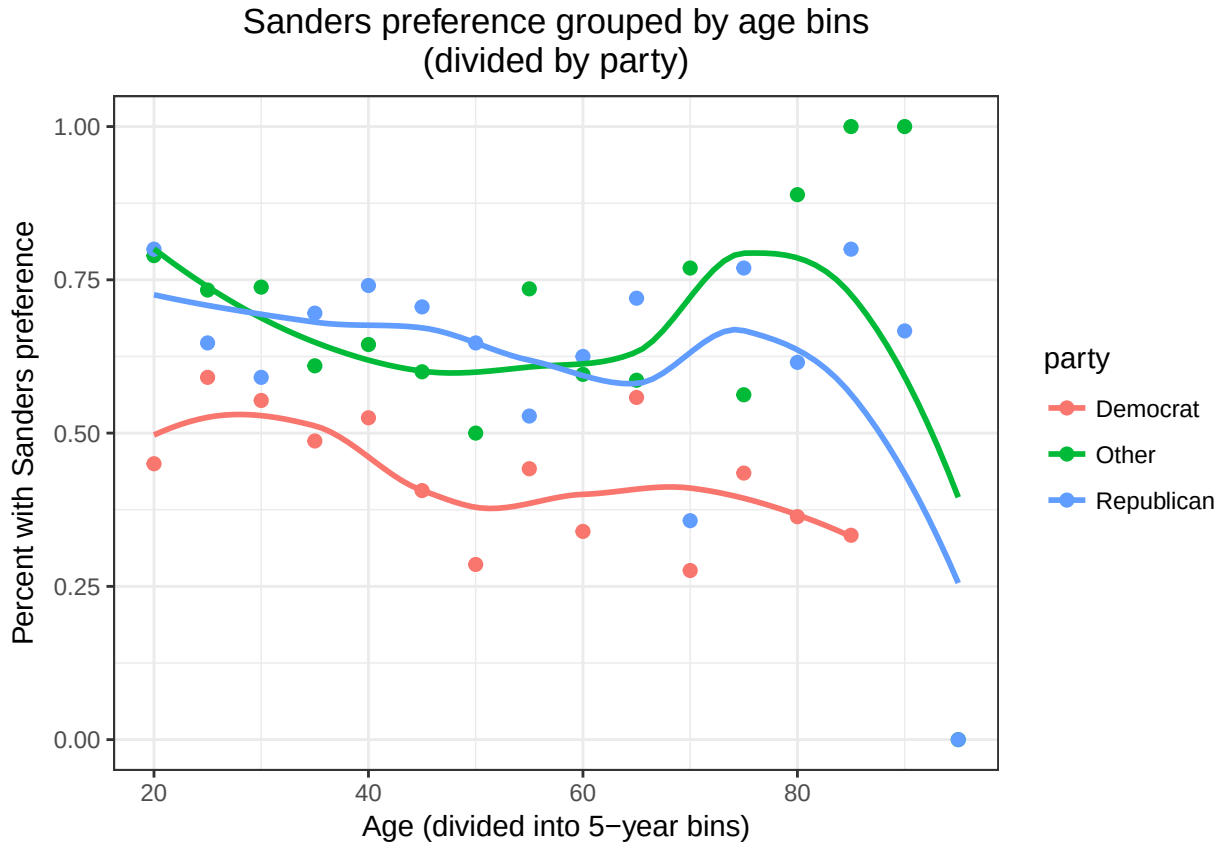

```
ggp <- ggplot(aes(x=as.factor(sanders_preference), y=age,
                  fill=as.factor(sanders_preference)),
              data=publicopinion_narm)

ggp+geom_violin()+geom_boxplot(width=0.33, fill="white")+
  facet_grid(partyfactor~racefactor)+
  scale_fill_discrete(name="Sanders Preference",
                      labels=c("No", "Yes"))+
  ggtitle("Age distribution split by Sanders preference\nseparated by race and party")+
  theme(plot.title = element_text(hjust = 0.5))
```



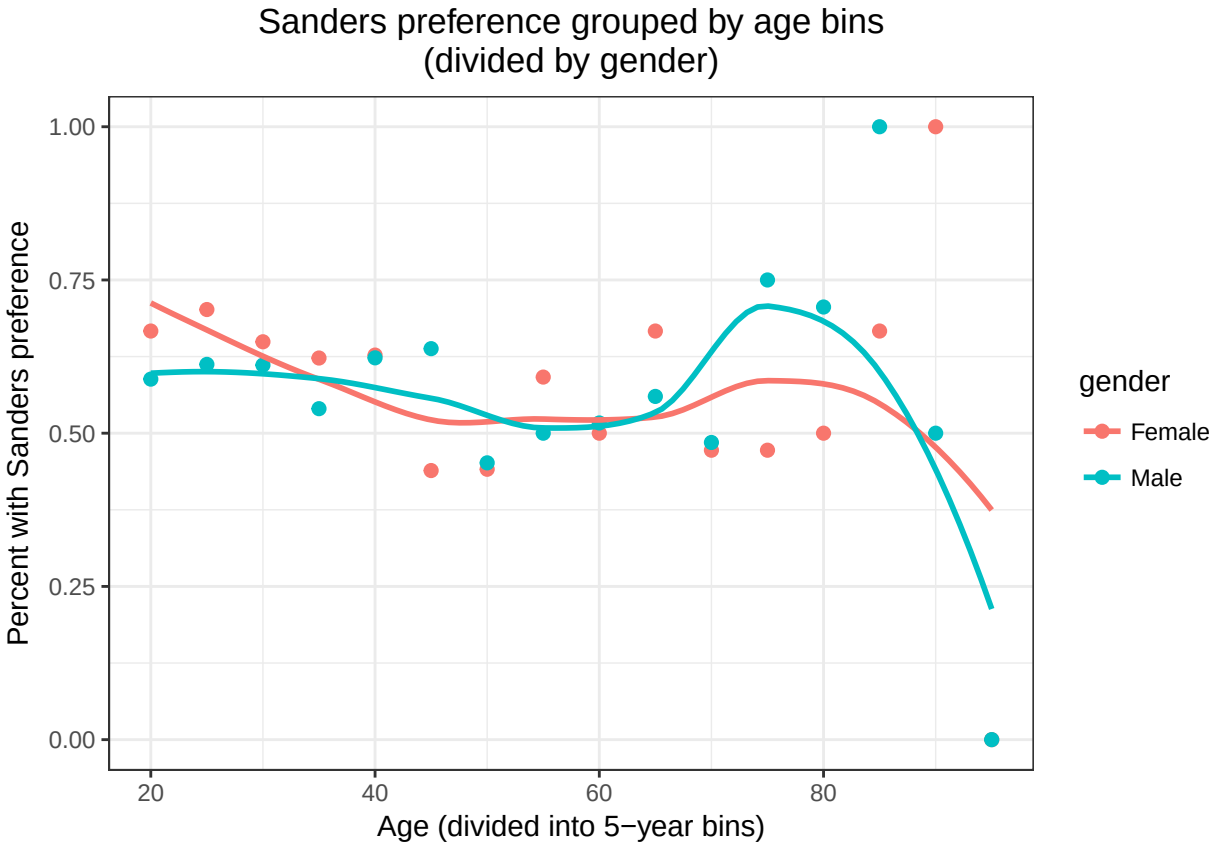

```
po_age_agg <- with(publicopinion_narm,
                   aggregate(age, list(gender=genderfactor,
                                       party=partyfactor,
                                       sanders_preference=sanders_preference,
                                       race=racefactor), mean))
dodge = position_dodge(width=0.9)
ggp <- ggplot(po_age_agg, aes(x=as.factor(sanders_preference), y=x, group=gender))
ggp +
  geom_bar(stat="identity", position=dodge, color="black", aes(fill=gender))+
  facet_wrap(race~party)+ylab("Age")+ggtitle("Mean age for all subgroups")+
  theme(plot.title=element_text(hjust=.5))
```



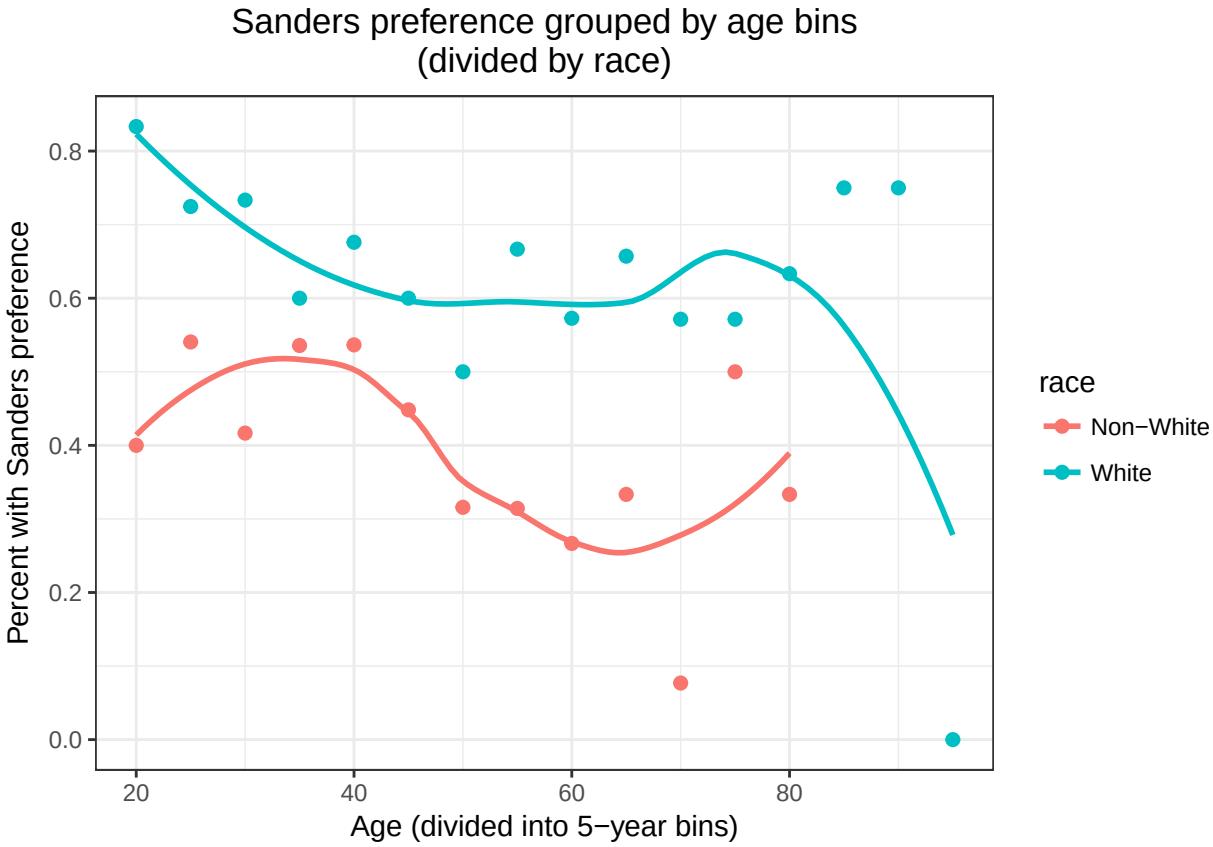

```
age_bin_agg_party <- with(publicopinion_narm,
                          aggregate(cbind(sanders_preference),
                                    list(agebin=5*round(age/5),
                                         party=partyfactor), mean))

ggp <- ggplot(age_bin_agg_party, aes(x=agebin, y=sanders_preference,
                                     color=party))

ggp + geom_point(aes(color=party), size=2)+
  geom_smooth(method="loess", se=F)+
  ylab("Percent with Sanders preference")+
  xlab("Age (divided into 5-year bins)")+
  ggtitle("Sanders preference grouped by age bins\n(divided by party)")+
  theme(plot.title=element_text(hjust=.5))
```




```
age_bin_agg_gender <- with(publicopinion_narm,
                           aggregate(cbind(sanders_preference),
                                     list(agebin=5*round(age/5),
                                          gender=genderfactor), mean))

ggp <- ggplot(age_bin_agg_gender, aes(x=agebin, y=sanders_preference,
                                      color=gender))

ggp + geom_point(aes(color=gender), size=2)+
  geom_smooth(method="loess", se=F)+
  ylab("Percent with Sanders preference")+
  xlab("Age (divided into 5-year bins)")+
  ggtitle("Sanders preference grouped by age bins\n(divided by gender)")+
  theme(plot.title=element_text(hjust=.5))
```




```
age_bin_agg_race <- with(publicopinion_narm,
                         aggregate(cbind(sanders_preference),
                                   list(agebin=5*round(age/5),
                                        race=racefactor), mean))

ggp <- ggplot(age_bin_agg_race, aes(x=agebin, y=sanders_preference,
                                    color=race))

ggp + geom_point(aes(color=race), size=2)+
  geom_smooth(method="loess", se=F)+
  ylab("Percent with Sanders preference")+
  xlab("Age (divided into 5-year bins)")+
  ggtitle("Sanders preference grouped by age bins\n(divided by race)")+
  theme(plot.title=element_text(hjust=.5))
```


Sanders preference grouped by age bins
(divided by race)
Percent with Sanders preference
0.0
0.2
0.4
0.6
0.8
20
40
60
80
Age (divided into 5−year bins)
race
Non−White
White