Giddy over ggplot

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.0.6      v dplyr   1.0.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
mpg
```

```
## # A tibble: 234 x 11
##    manufacturer model    displ  year   cyl trans   drv     cty   hwy fl    class
##    <chr>        <chr>    <dbl> <int> <int> <chr>   <chr> <int> <int> <chr> <chr>
##  1 audi         a4         1.8  1999     4 auto(l~ f        18    29 p     comp~
##  2 audi         a4         1.8  1999     4 manual~ f        21    29 p     comp~
##  3 audi         a4         2    2008     4 manual~ f        20    31 p     comp~
##  4 audi         a4         2    2008     4 auto(a~ f        21    30 p     comp~
##  5 audi         a4         2.8  1999     6 auto(l~ f        16    26 p     comp~
##  6 audi         a4         2.8  1999     6 manual~ f        18    26 p     comp~
##  7 audi         a4         3.1  2008     6 auto(a~ f        18    27 p     comp~
##  8 audi         a4 quat~   1.8  1999     4 manual~ 4        18    26 p     comp~
##  9 audi         a4 quat~   1.8  1999     4 auto(l~ 4        16    25 p     comp~
## 10 audi         a4 quat~   2    2008     4 manual~ 4        20    28 p     comp~
## # ... with 224 more rows
```
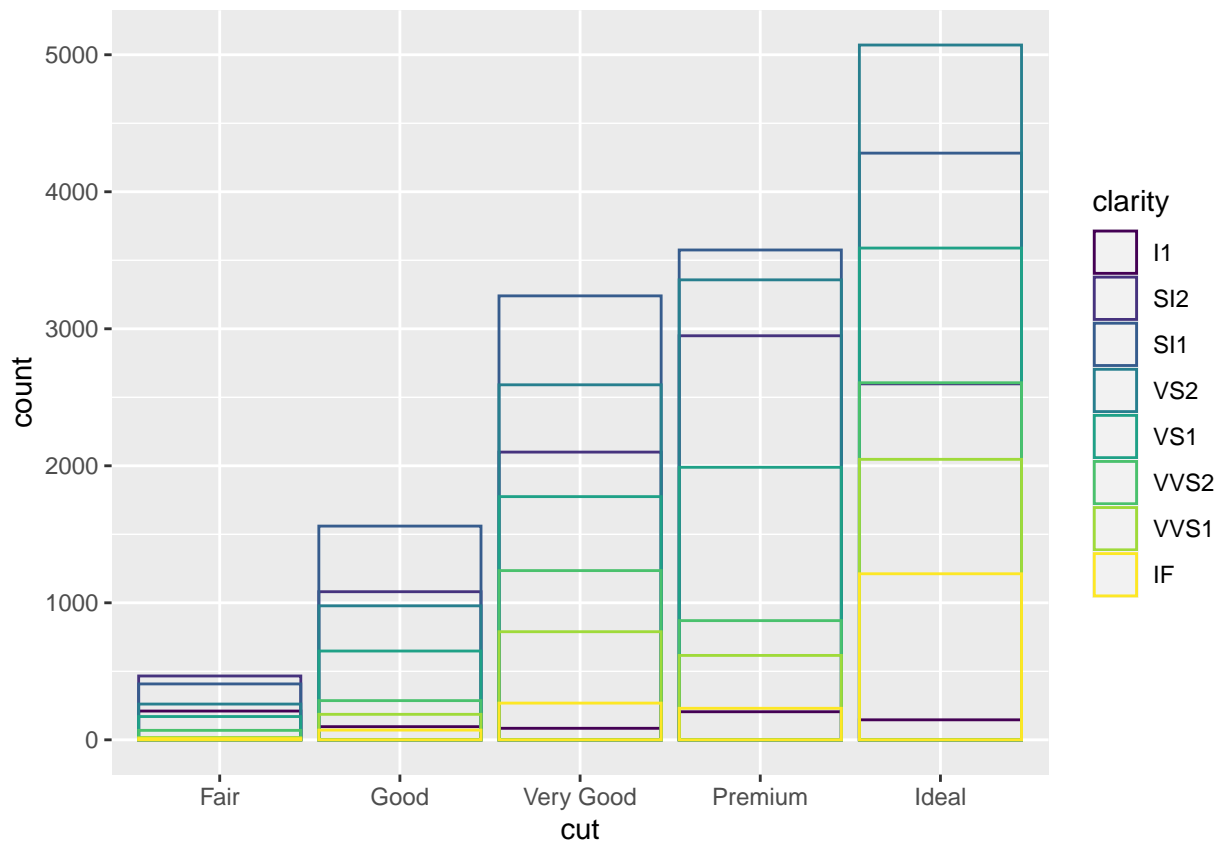
```
mpg%>%mutate(cyl=as.character(cyl))%>%
  ggplot(mapping=aes(x=displ,y=hwy,shape=cyl,color=class))+
  geom_point()+
  geom_jitter()
```
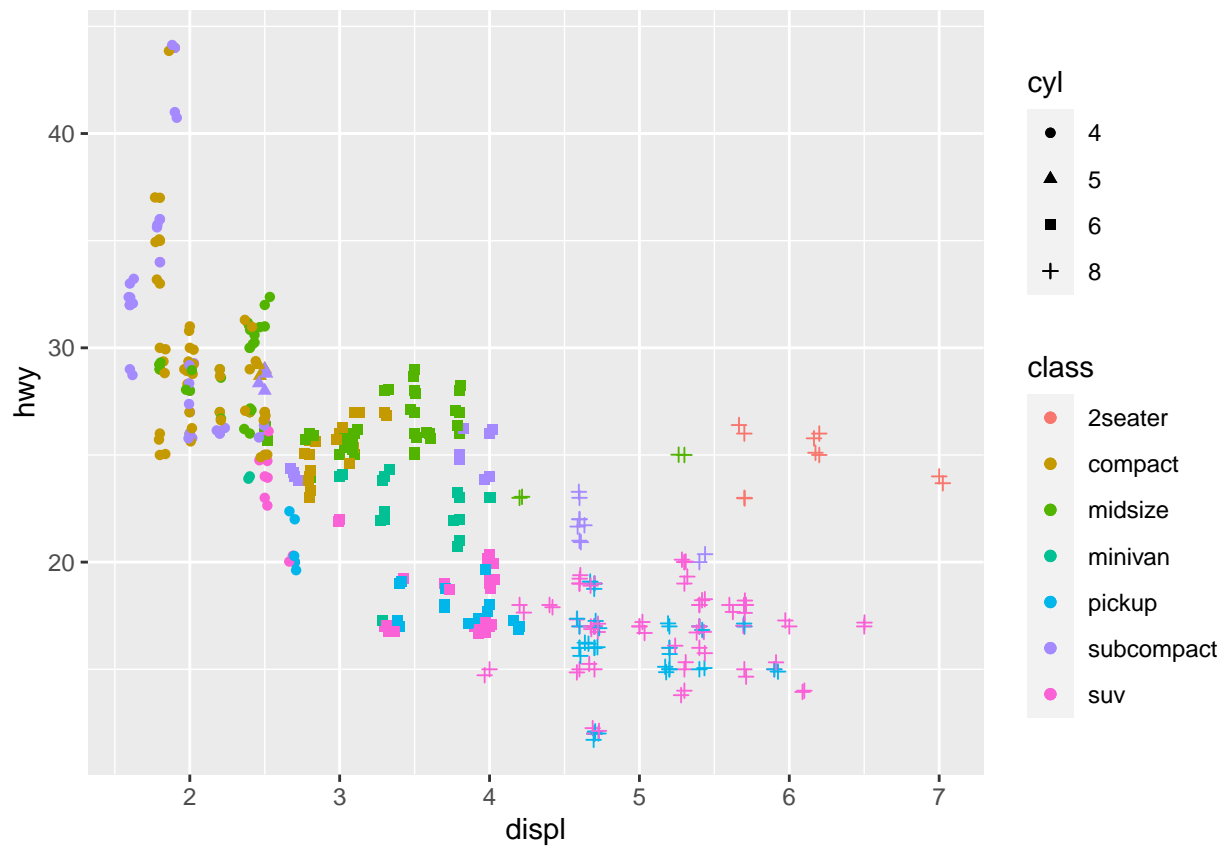
```
diamonds%>%ggplot()+
  geom_bar(mapping=aes(x=cut,color=clarity),position="identity",alpha=.3,fill=NA)
```

```
gss_cat
```

```
## # A tibble: 21,483 x 9
##     year marital     age race  rincome     partyid      relig      denom     tvhours
##    <int> <fct>     <int> <fct> <fct>       <fct>        <fct>      <fct>        <int>
##  1  2000 Never ma~    26 White $8000 to ~  Ind,near r~  Protesta~  Souther~        12
##  2  2000 Divorced     48 White $8000 to ~  Not str re~  Protesta~  Baptist~        NA
##  3  2000 Widowed      67 White Not appli~  Independent  Protesta~  No deno~         2
##  4  2000 Never ma~    39 White Not appli~  Ind,near r~  Orthodox~  Not app~         4
##  5  2000 Divorced     25 White Not appli~  Not str de~  None       Not app~         1
##  6  2000 Married      25 White $20000 - ~  Strong dem~  Protesta~  Souther~        NA
##  7  2000 Never ma~    36 White $25000 or~  Not str re~  Christian  Not app~         3
##  8  2000 Divorced     44 White $7000 to ~  Ind,near d~  Protesta~  Luthera~        NA
##  9  2000 Married      44 White $25000 or~  Not str de~  Protesta~  Other            0
## 10  2000 Married      47 White $25000 or~  Strong rep~  Protesta~  Souther~         3
## # ... with 21,473 more rows
```

```
gss_cat%>%
  count(age)%>%left_join(gss_cat)%>%
  filter(partyid !="Don't know", partyid !="No answer", partyid != "Other party", partyid != "Independen
  mutate(Party= factor(partyid))%>%
  mutate(Party= fct_collapse(partyid, "Ind" = c("Other party", "Independent")))%>%
  mutate(Party= fct_relevel(Party, "Strong democrat", "Not str democrat","Ind, near dem", "Ind","Ind,nea
  mutate(Party=as.numeric(Party))%>%
  filter(!is.na(age))%>%
  group_by(age)%>%
```

```
  summarize(Party=mean(Party),avg=mean(tvhours,na.rm=TRUE),Number=n)%>%
  ggplot(mapping=aes(x=age, y=avg))+
  geom_point(mapping=aes(color=Party,size=Number))+
  scale_color_gradient(high="red",low="blue")+
  geom_smooth(color="black")+
  labs(x="Age",y="Hours",title="Average TV hours per day by age and political affiliation")
```
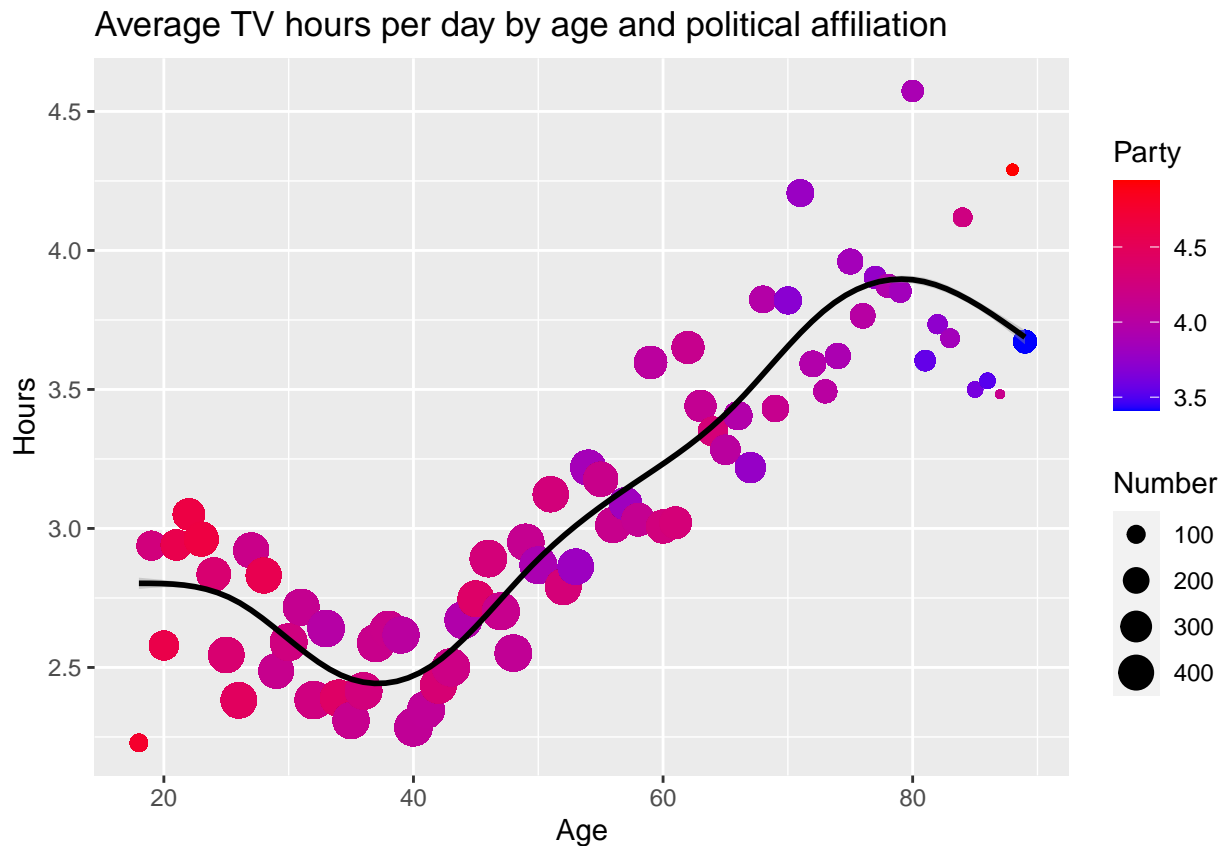
```
## Joining, by = "age"
```

```
## Warning: Unknown levels in 'f': Ind, near dem
```

```
## 'summarise()' has grouped output by 'age'. You can override using the '.groups' argument.
```

```
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```
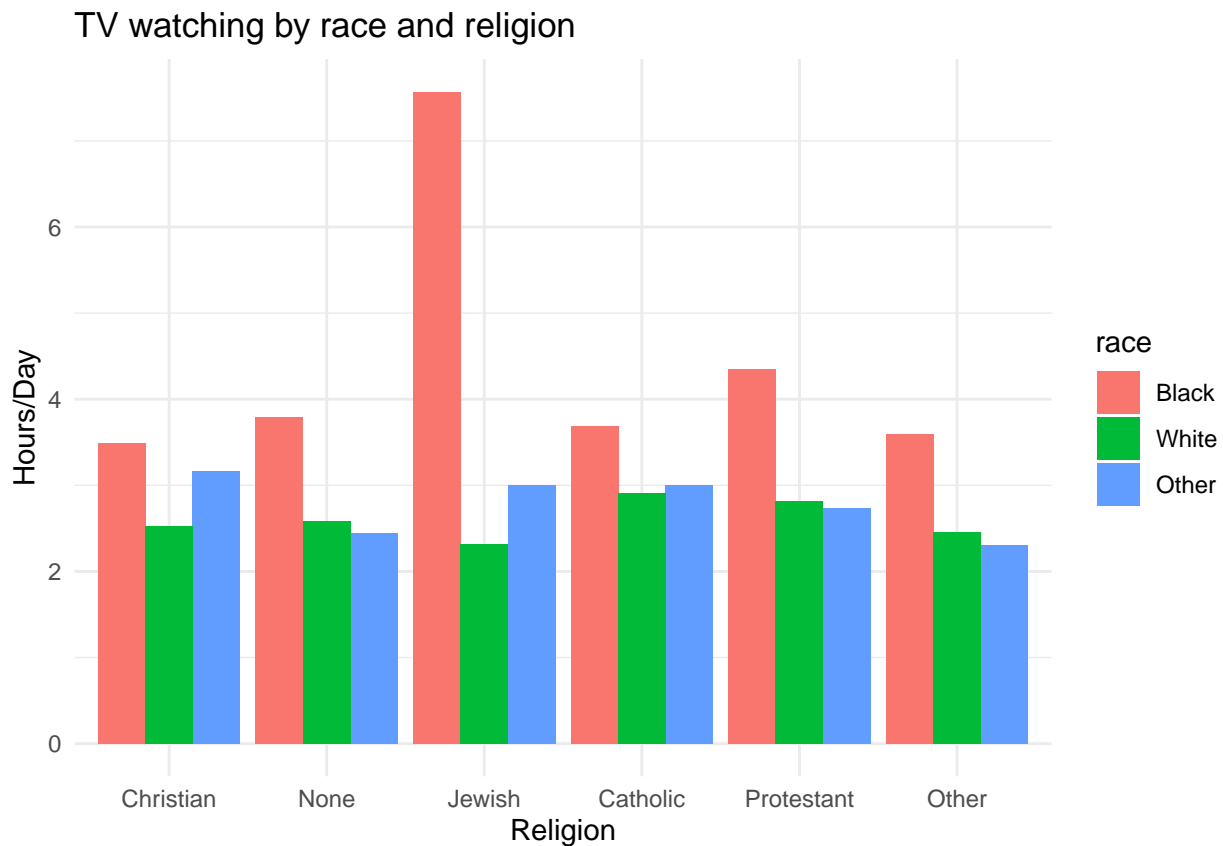


```
tvwatching_byraceandreligion <- gss_cat%>%
  mutate(relig = fct_lump(relig, n = 5)) %>%
  group_by(race,relig)%>%
  summarize(avg_tvhours=mean(tvhours, na.rm= TRUE))%>%
  mutate(race=fct_relevel(race,"Black","White","Other"))%>%
  ggplot()+
  geom_col(mapping=aes(x=relig,y=avg_tvhours,fill = race), position= "dodge")+
  labs(x="Religion",y="Hours/Day", title="TV watching by race and religion")+
  theme_minimal()
```
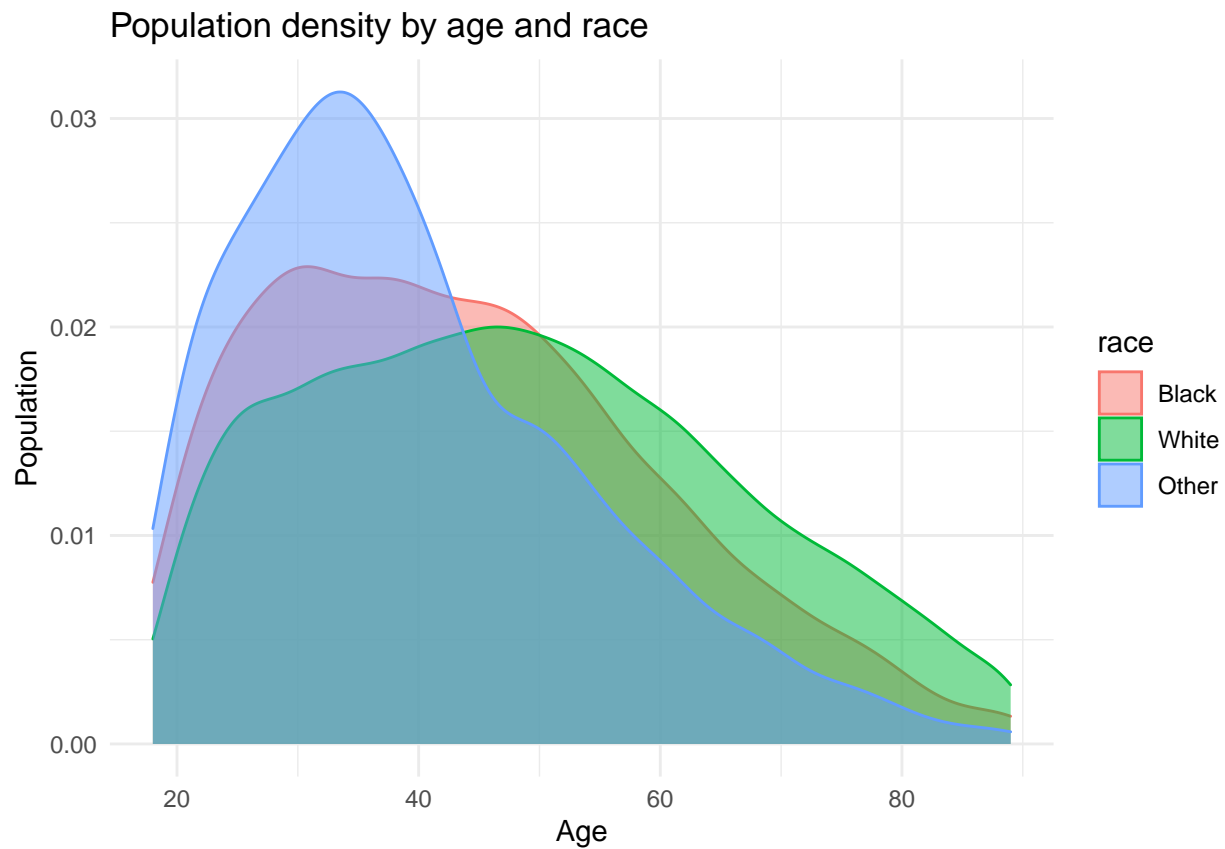
```
tvwatching_byraceandreligion
```

TV watching by race and religion


```
gss_cat%>%
  mutate(race=fct_relevel(race, "Black", "White", "Other"))%>%
  ggplot()+
  geom_density(aes(age,fill=race, color=race),alpha=0.5)+
  labs(x="Age",y="Population", title="Population density by age and race")+
  theme_minimal()
```

Population density by age and race


```
gss_cat%>%
  filter(relig !="Don't know", relig !="No answer", relig !="Other")%>%
  count(relig)%>%
  mutate(relig=fct_reorder(relig, n))%>%
  ggplot()+
  geom_bar(aes(x = "relig", y = n, fill = relig),
           stat = "identity", width = 0.4) +
  coord_polar("y")+
  labs(title="GSS Religions")+
  theme_void()
```

GSS Religions

relig
- Native american
- Other eastern
- Hinduism
- Orthodox−christian
- Moslem/islam
- Inter−nondenominational
- Buddhism
- Jewish
- Christian
- None
- Catholic
- Protestant