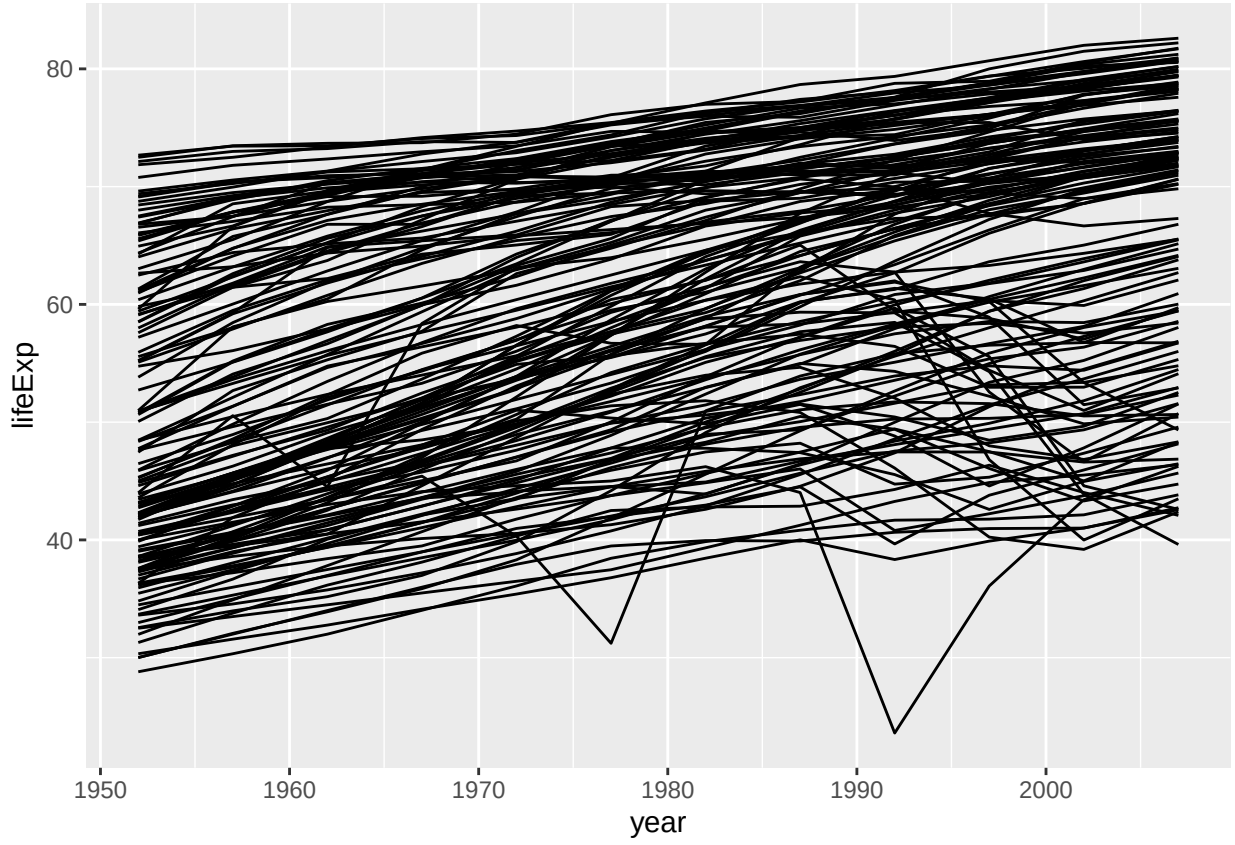
194161016_Lab2

```
library(gapminder)
library(ggplot2)
data.gapminder <- data.frame(gapminder)
colnames(data.gapminder)
```

```
## [1] "country"   "continent" "year"      "lifeExp"   "pop"       "gdpPercap"
```

```
ggplot(data = data.gapminder)+
  geom_line(mapping = aes(x=year,y=lifeExp,group=country))
```

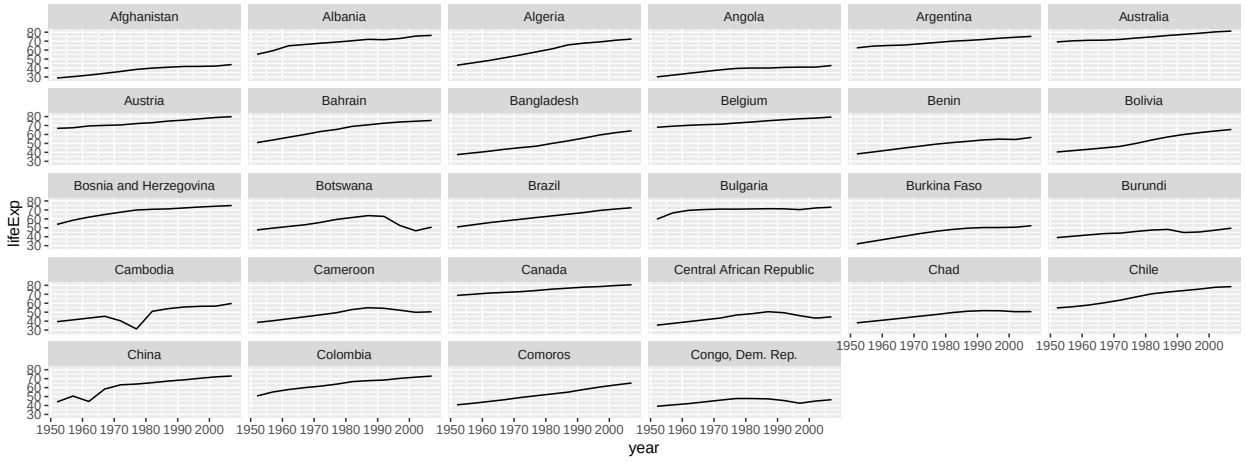


1)The graph looks a bit clumsy as we have 142 countries.
2)The plot has many lines, where each line giving us life expectancy over time for every country.The plot is not that meaningful as we may not be able to figure out which line corresponds to which country.

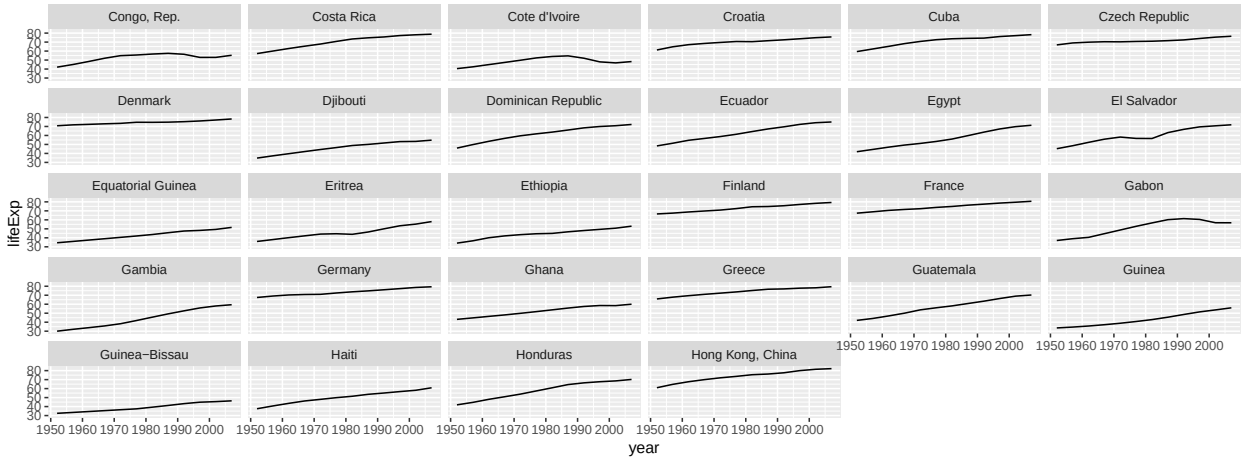
```
#2.1
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

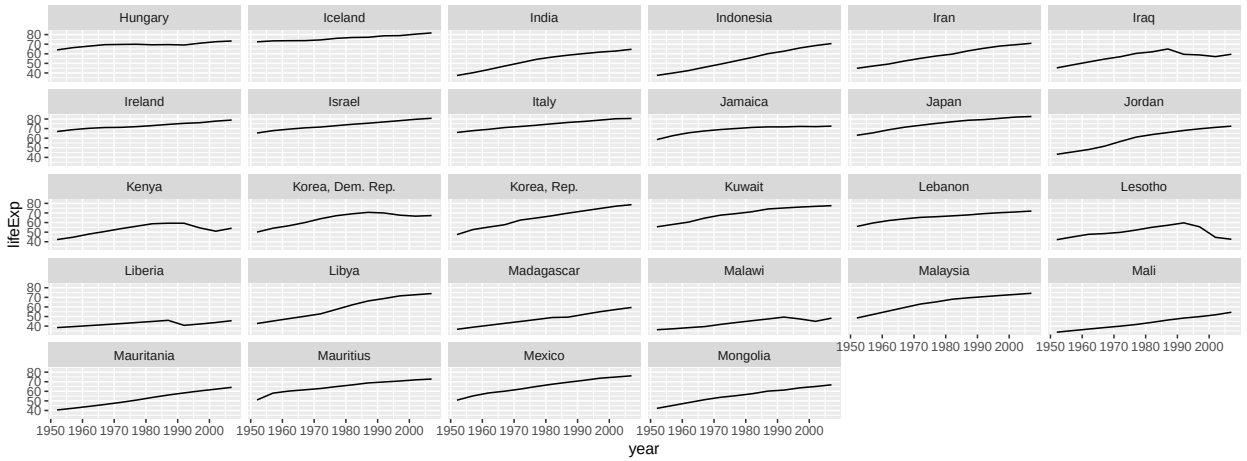
```
library(socviz)
countries <- unique(data.gapminder$country)
data.g1 <- filter(data.gapminder,data.gapminder$country %in% countries[1:28])
data.g2 <- filter(data.gapminder,data.gapminder$country %in% countries[29:56])
data.g3 <- filter(data.gapminder,data.gapminder$country %in% countries[57:84])
data.g4 <- filter(data.gapminder,data.gapminder$country %in% countries[85:112])
data.g5 <- filter(data.gapminder,data.gapminder$country %in% countries[113:142])



ggplot(data = data.g1)+
  geom_line(mapping = aes(x=year,y=lifeExp))+
  facet_wrap(.~country)
```



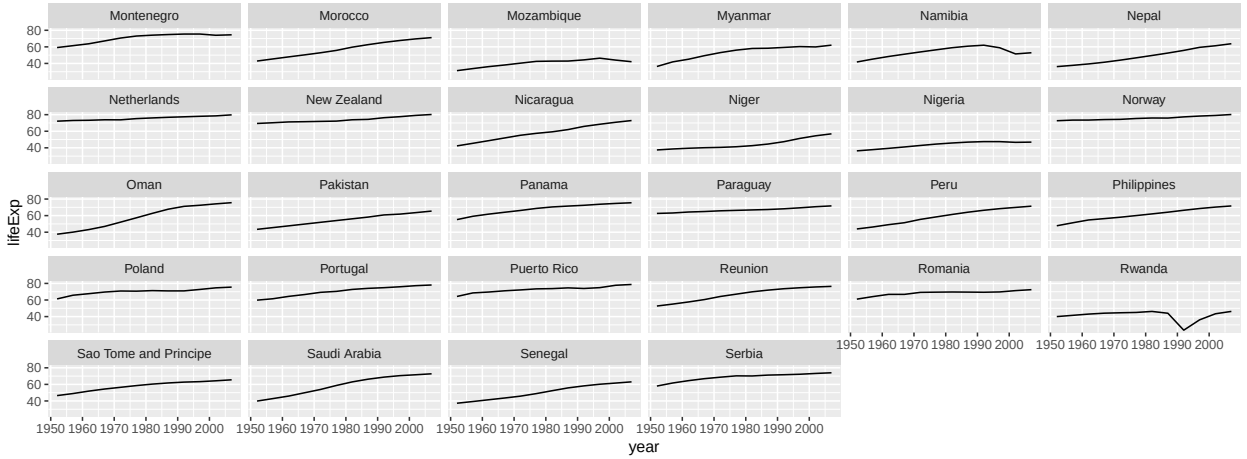

```
ggplot(data = data.g2)+
  geom_line(mapping = aes(x=year,y=lifeExp))+
  facet_wrap(.~country)
```



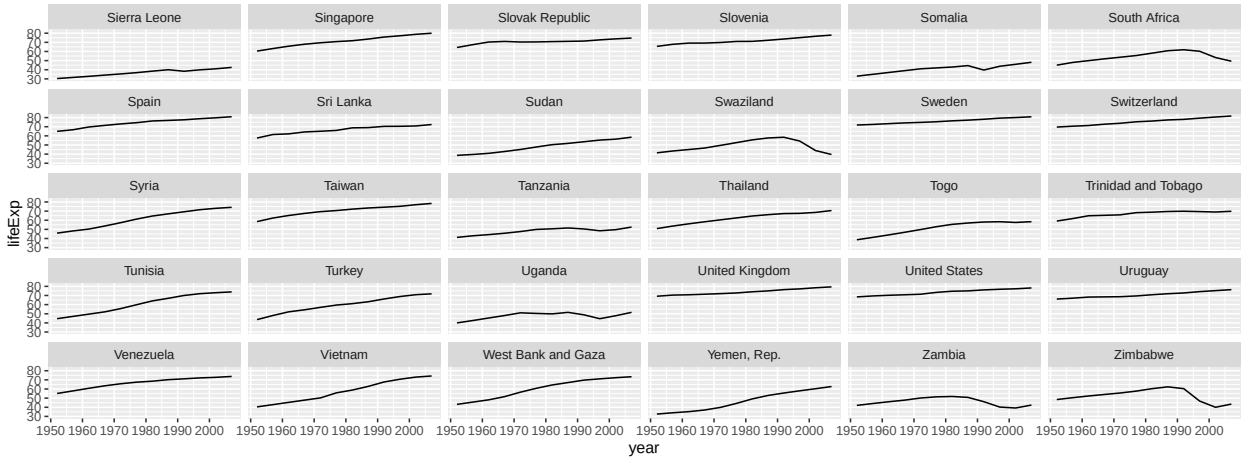

```
ggplot(data = data.g3)+
  geom_line(mapping = aes(x=year,y=lifeExp))+
  facet_wrap(.~country)
```




```
ggplot(data = data.g4)+
  geom_line(mapping = aes(x=year,y=lifeExp))+
  facet_wrap(.~country)
```



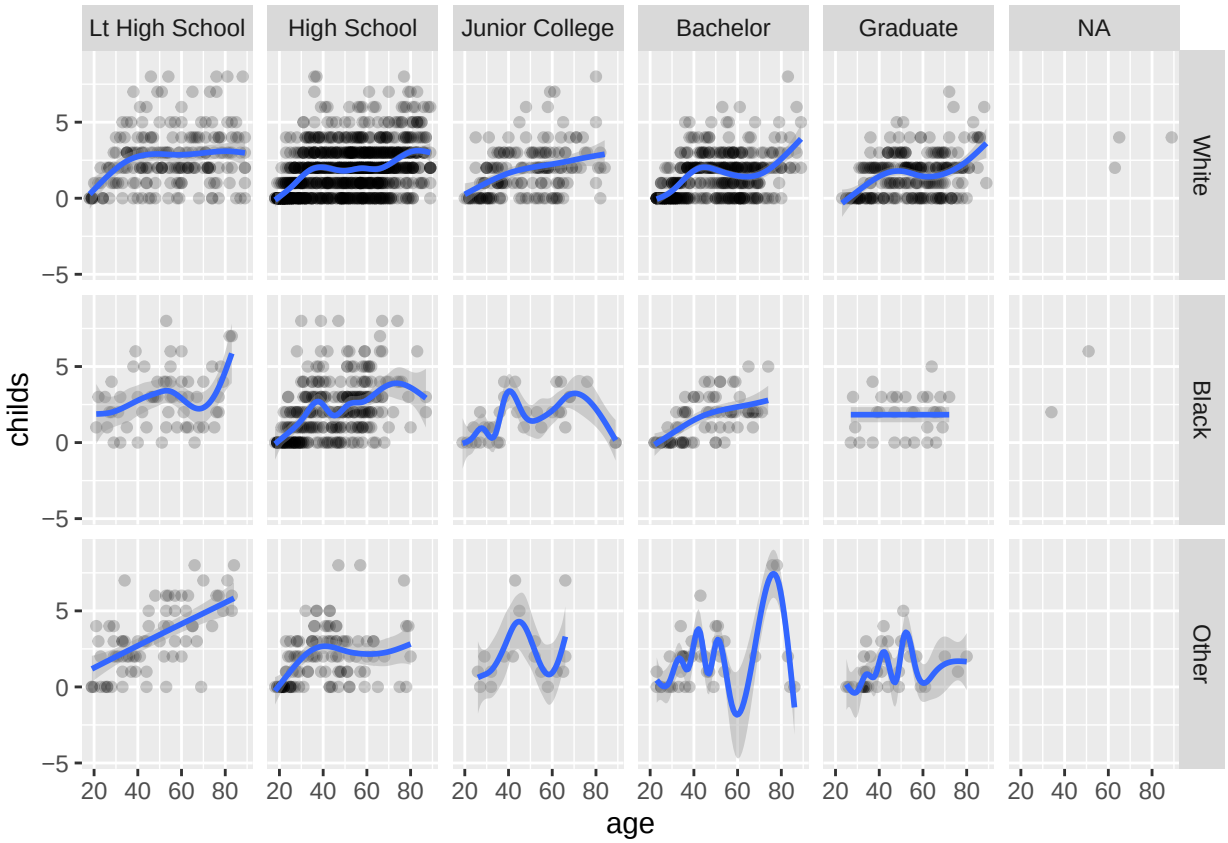
```
ggplot(data = data.g5)+
  geom_line(mapping = aes(x=year,y=lifeExp))+
  facet_wrap(.~country)
```



1) Country is the most appropriate variable to facet the data.
2) As we can see in the plot that upon using country to facet the data,we can clearly observe the trends of life expectancy over time for every country.

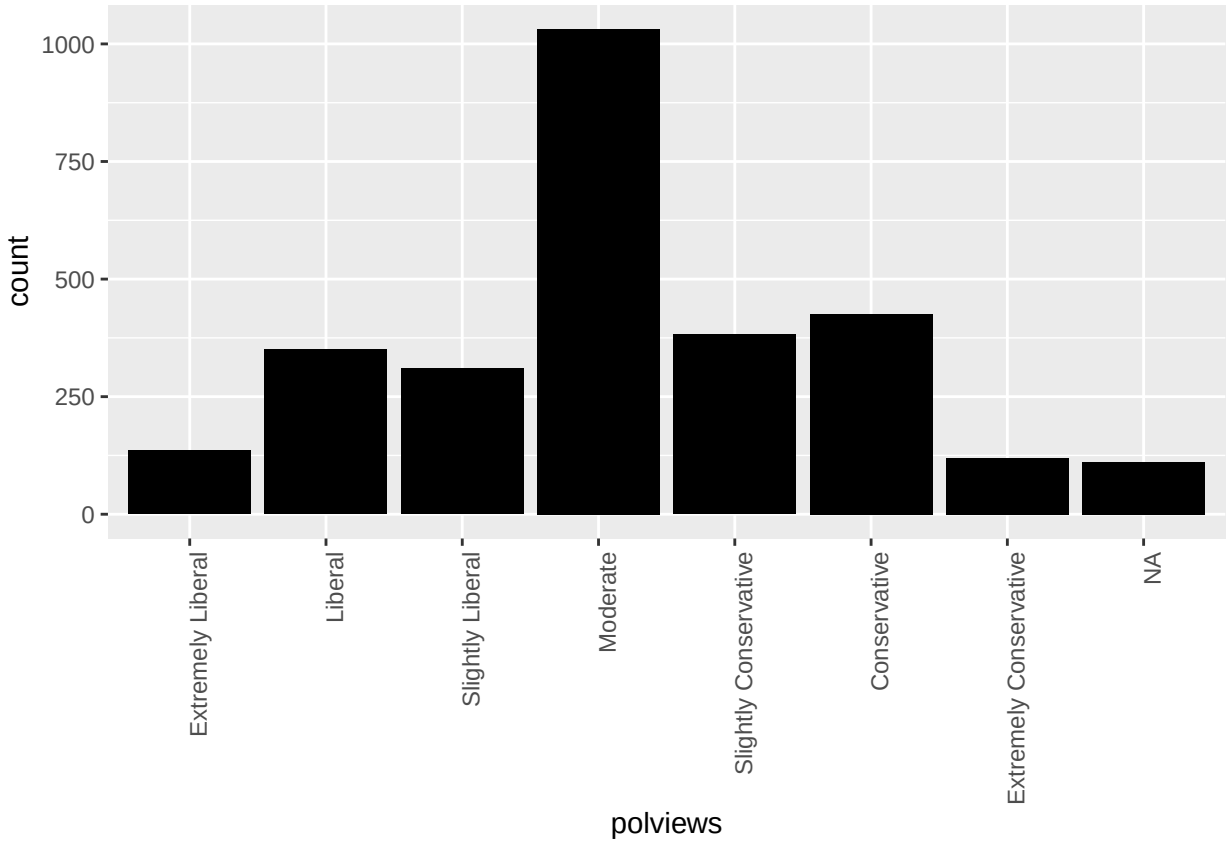
```
data.gss <- data.frame(gss_sm)

#2.2
ggplot(data=data.gss,mapping=aes(x=age,y=childs))+geom_point(alpha=0.2)+geom_smooth()+facet_grid(race~d
```

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

## Warning: Removed 18 rows containing non-finite values (stat_smooth).

## Warning: Computation failed in `stat_smooth()`:
## x has insufficient unique values to support 10 knots: reduce k.

## Warning: Computation failed in `stat_smooth()`:
## x has insufficient unique values to support 10 knots: reduce k.

## Warning: Removed 18 rows containing missing values (geom_point).
```

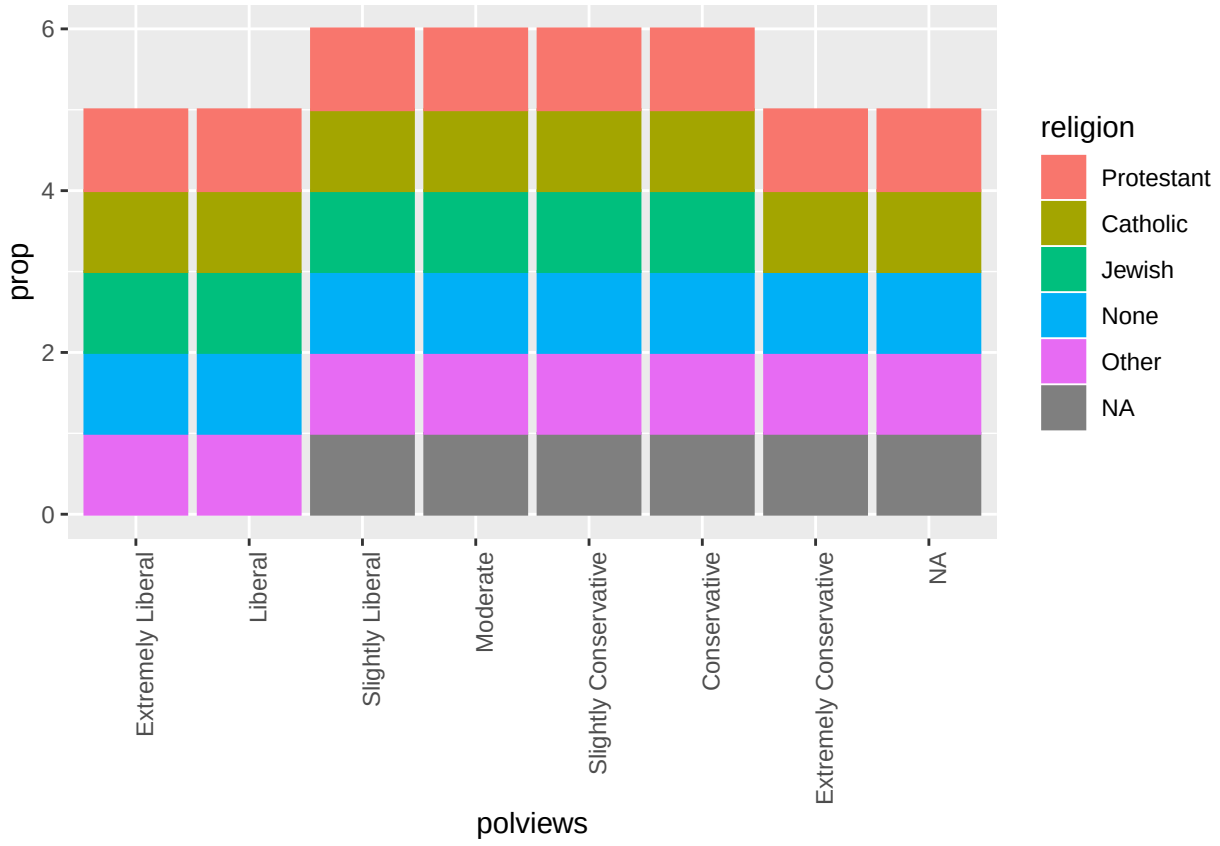


1) The above smooth plots tells us age and number of childern with that age for different combinations of race and degree.
2) We can see White and High school category has more number of children where as there is no child from white, balck or other with no degree.

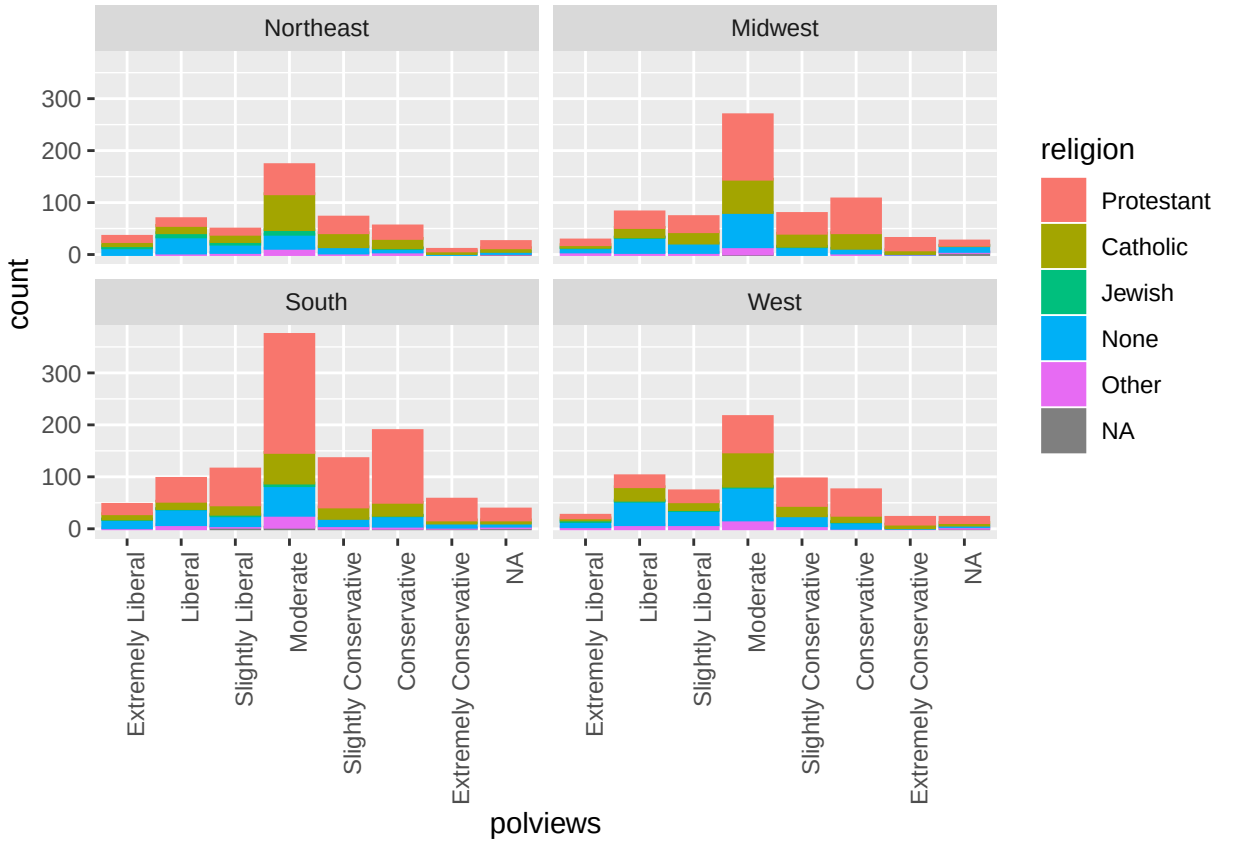
```
#3
ggplot(data=data.gss)+
  geom_bar(mapping=aes(x=polviews),fill='black')+theme(axis.text.x = element_text(angle = 90, hjust = 1
```




```
#4
ggplot(data=data.gss)+
  geom_bar(mapping=aes(x=polviews,y=..prop..,fill=religion,color=religion))+theme(axis.text.x = element_
```




```
ggplot(data=data.gss)+
  geom_bar(mapping=aes(x=polviews,fill=religion,color=religion))+
  facet_wrap(.~bigregion)+theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



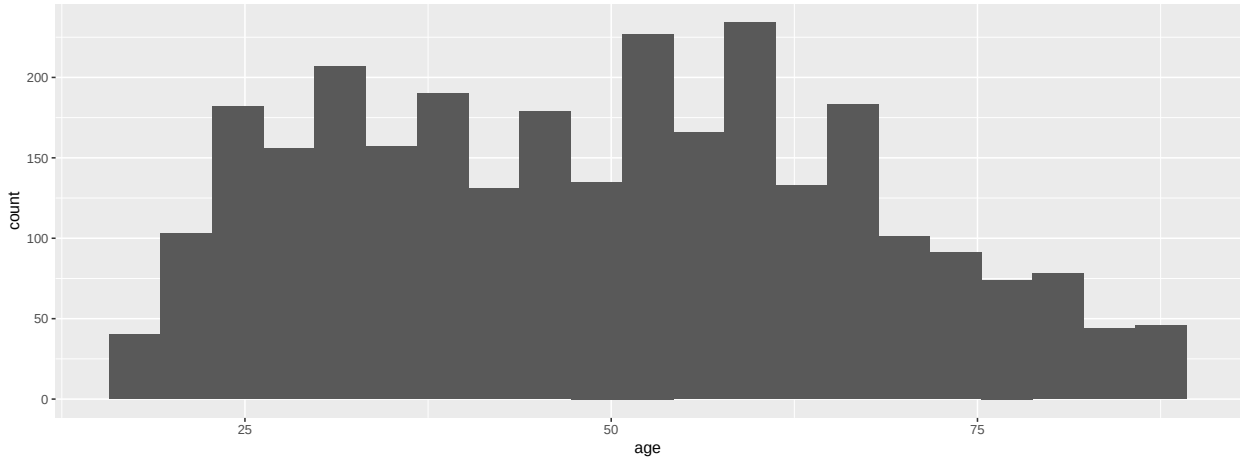

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

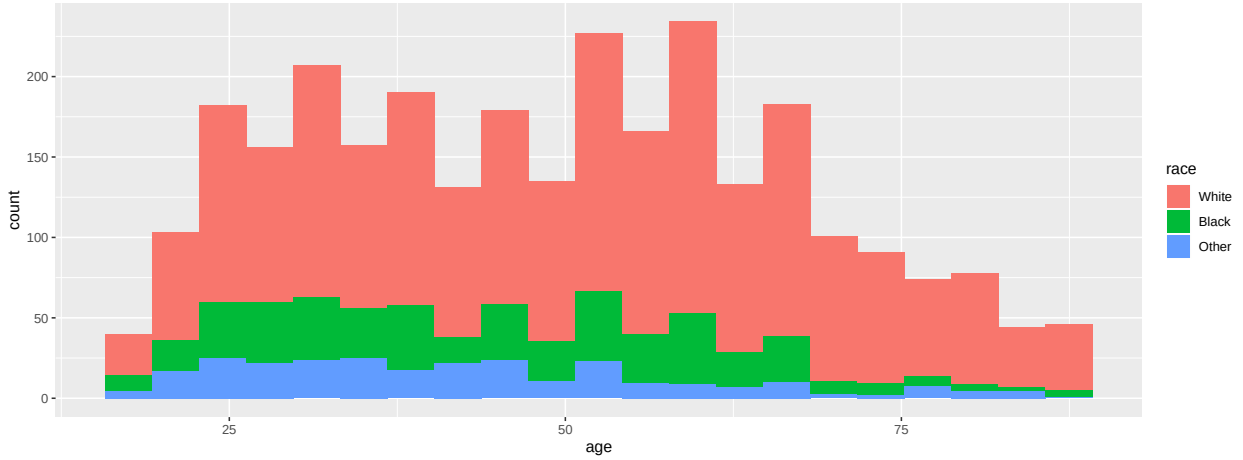
```
#5.1
ggplot(data=data.gss)+
  geom_histogram(mapping=aes(x=age),binwidth = 3.5)
```

```
## Warning: Removed 10 rows containing non-finite values (stat_bin).
```




```
#5.2
ggplot(data=data.gss)+
  geom_histogram(mapping=aes(x=age,fill=race),binwidth = 3.5)
```

```
## Warning: Removed 10 rows containing non-finite values (stat_bin).
```




```
#5.3

p1 <- ggplot(data=data.gss,mapping=aes(x=age,fill=race,color=race))+
  geom_density(position="stack",alpha=0.8)
p2<-ggplot(data=data.gss,mapping=aes(x=age,fill=race,color=race))+
  geom_histogram()

p3<-ggplot(data=data.gss,mapping=aes(x=agegrp,fill=race,color=race))+
  geom_density(position="stack",alpha=0.8)
p4<-ggplot(data=data.gss,mapping=aes(x=agegrp,fill=race,color=race))+
  geom_histogram(stat="count")
```

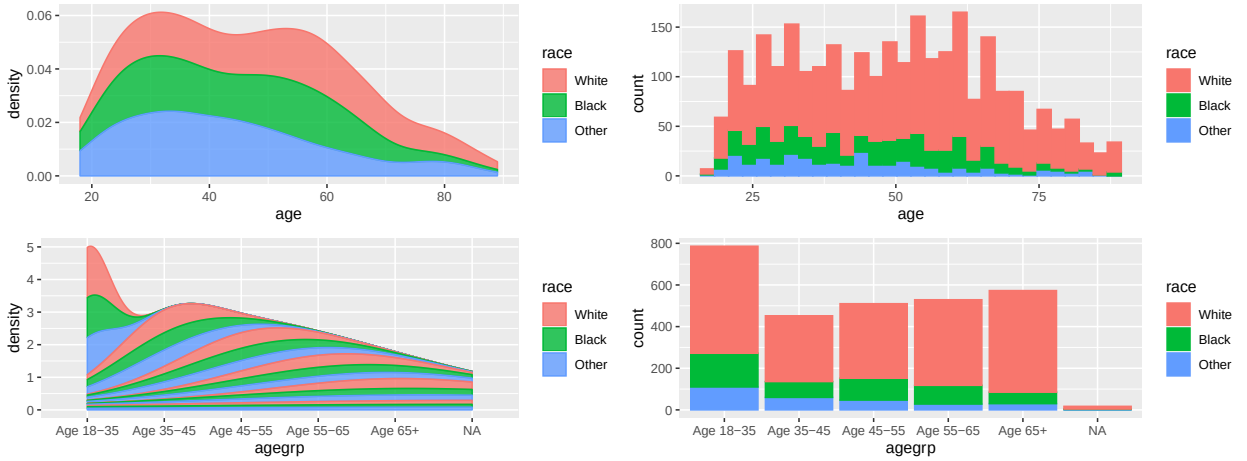
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
grid.arrange(p1,p2,p3,p4)
```

```
## Warning: Removed 10 rows containing non-finite values (stat_density).
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 10 rows containing non-finite values (stat_bin).
```



As age variable has continuous values, density plot as well as histogram will give us a good understanding. As agegrp variable has discrete values density plot will not give us much information as density best suits for continuous variable.where as histogram helps us with better understanding of the data.
If we look at the plots for agegrp density plot is completely overlapping and not giving proper information, where as histogram is clearly telling us count of people with different races for different age groups.Hence Histogram is useful for agegrp.

```
#5.4
ggplot(data=data.gss)+
  geom_density(mapping=aes(x=age,fill=degree,color=degree),alpha=0.3)+facet_wrap(.~income16,scale="free
```

```
## Warning: Removed 10 rows containing non-finite values (stat_density).
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.
```