Beautifying the Visualization

Bing-Je_Wu

8/17/2019

Outlines

- Modifying the Background
- Modifying Axes
- Modifying Scales
- Modifying Legends
- Annotating Plots
- Titles
- Themes

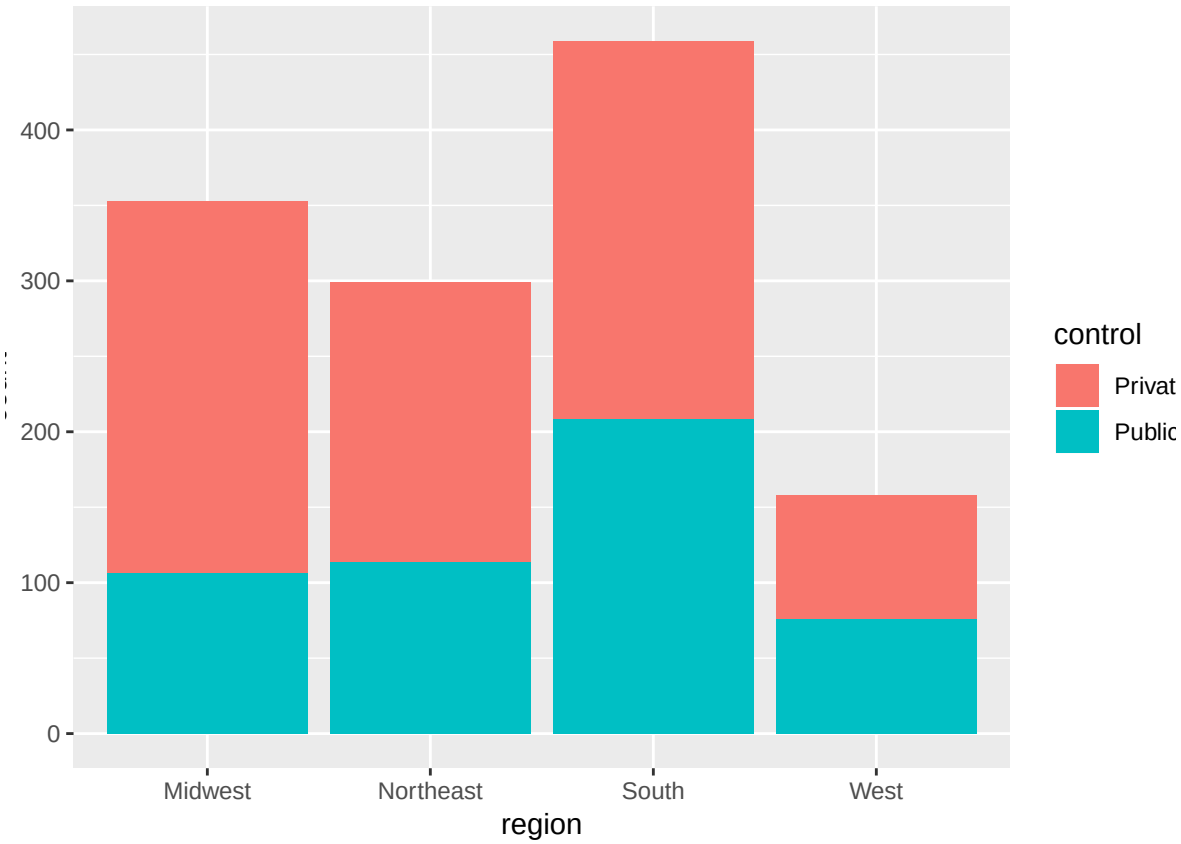
Modifying the Background

Load the dataset

```
library(tidyverse)
college <- read_csv('http://672258.youcanlearnit.net/college.csv')
college <- college %>%
  mutate(state=as.factor(state), region=as.factor(region),
         highest_degree=as.factor(highest_degree),
         control=as.factor(control), gender=as.factor(gender),
         loan_default_rate=as.numeric(loan_default_rate))
```

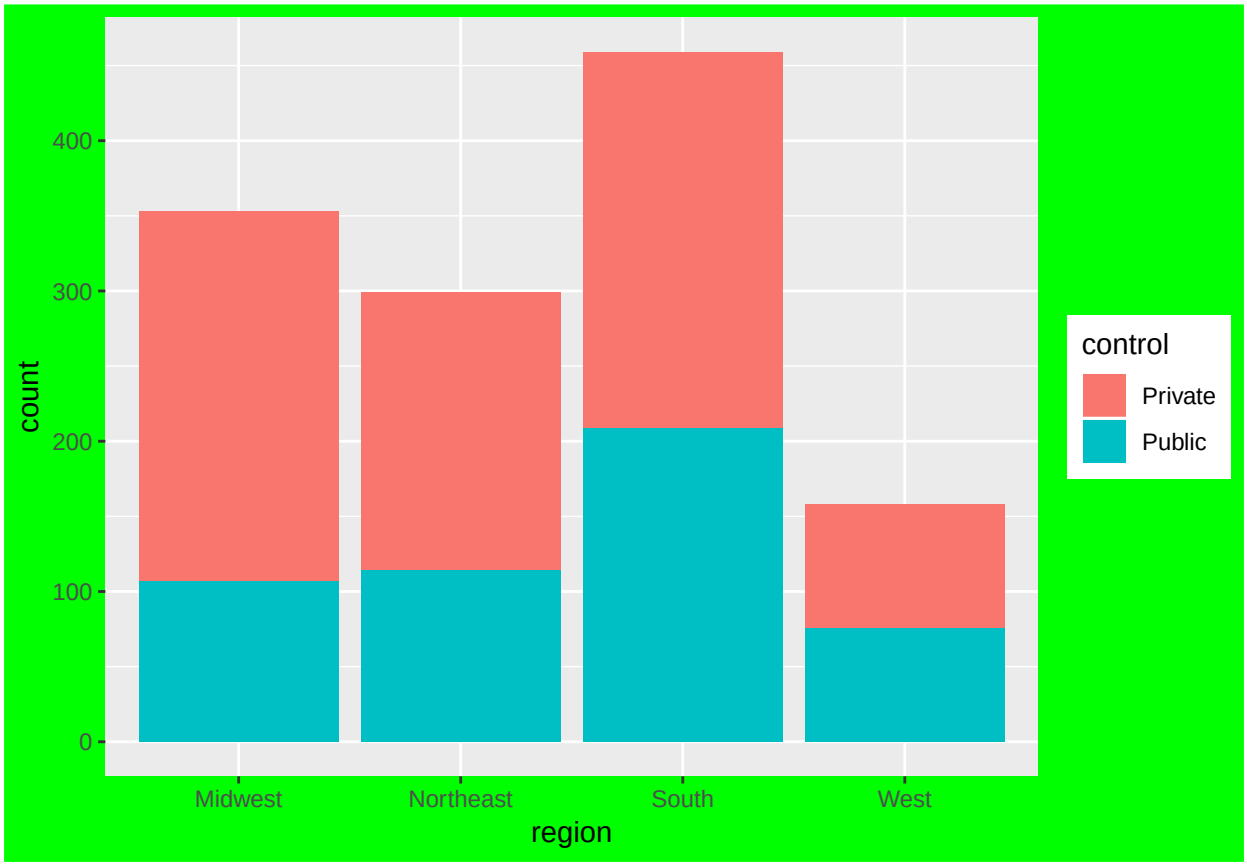
Create the bar graph

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control))
```



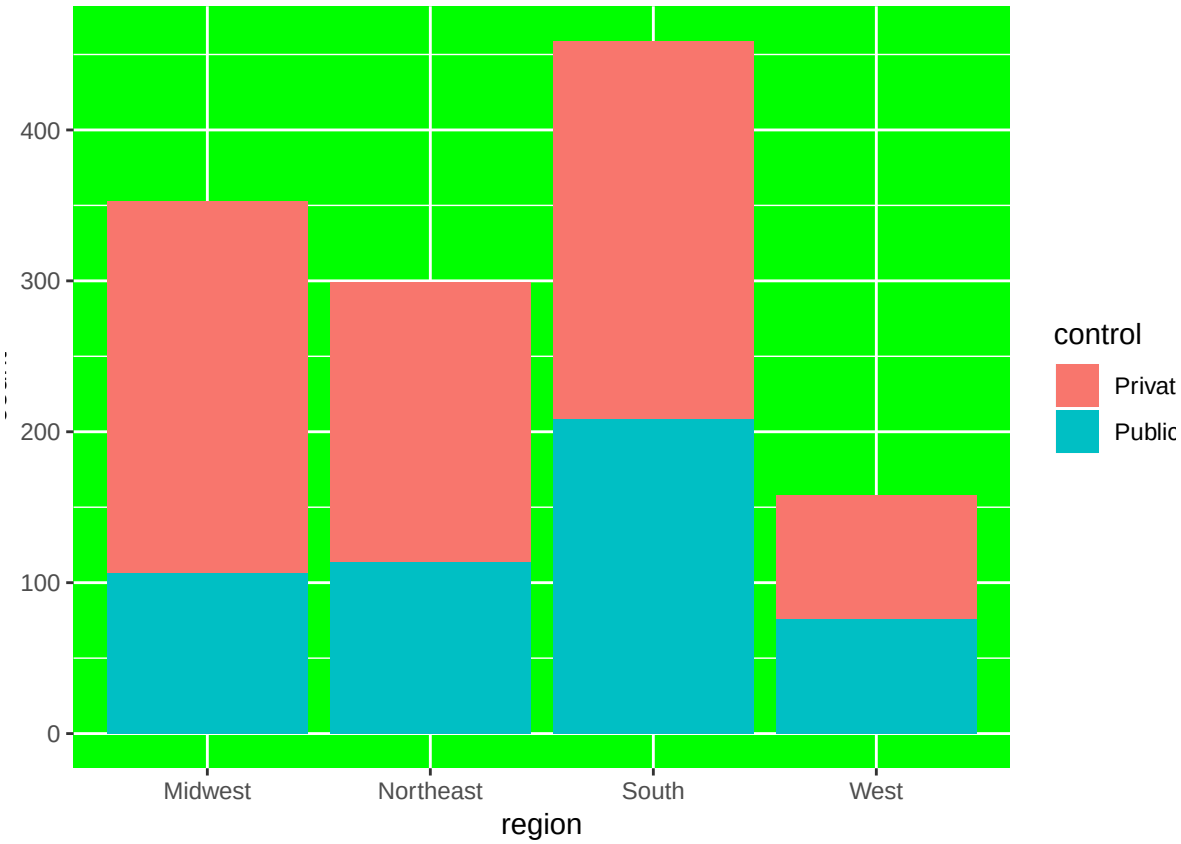
Change the "plot" background color

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(plot.background=element_rect(fill='green'))
```



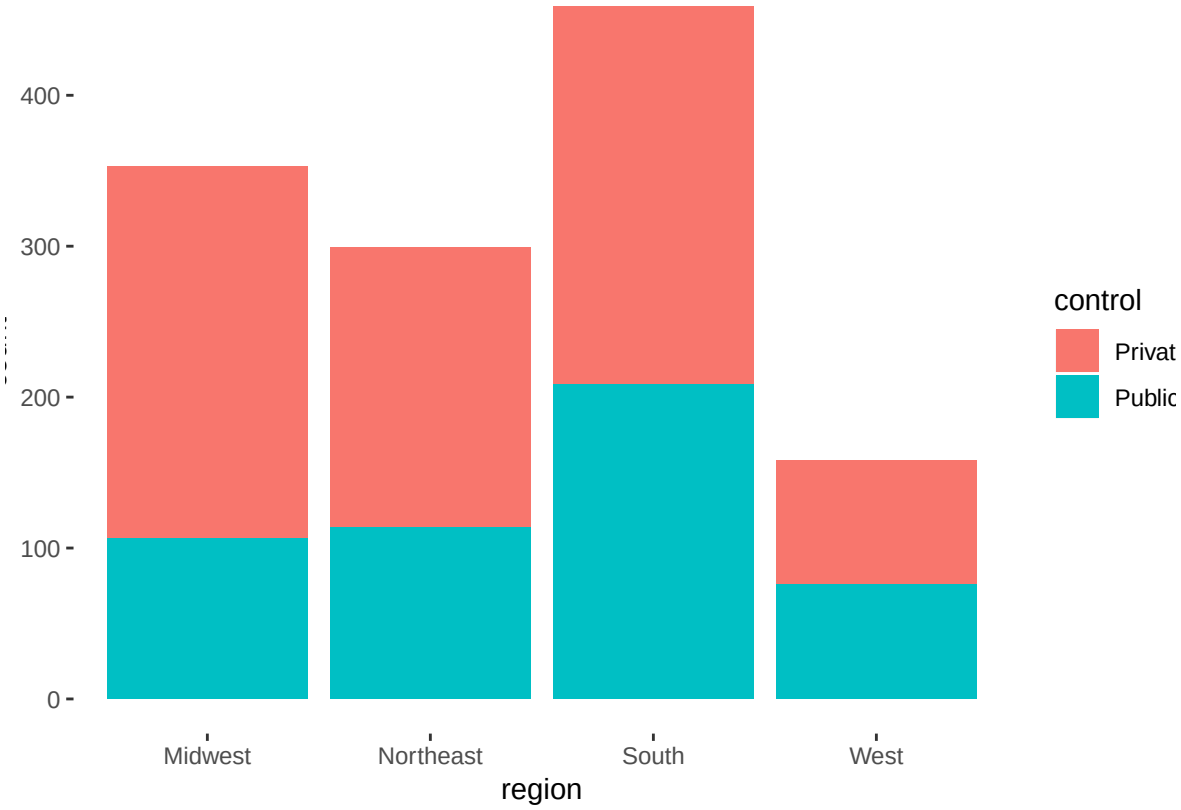
Change the "panel" background color

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_rect(fill='green'))
```



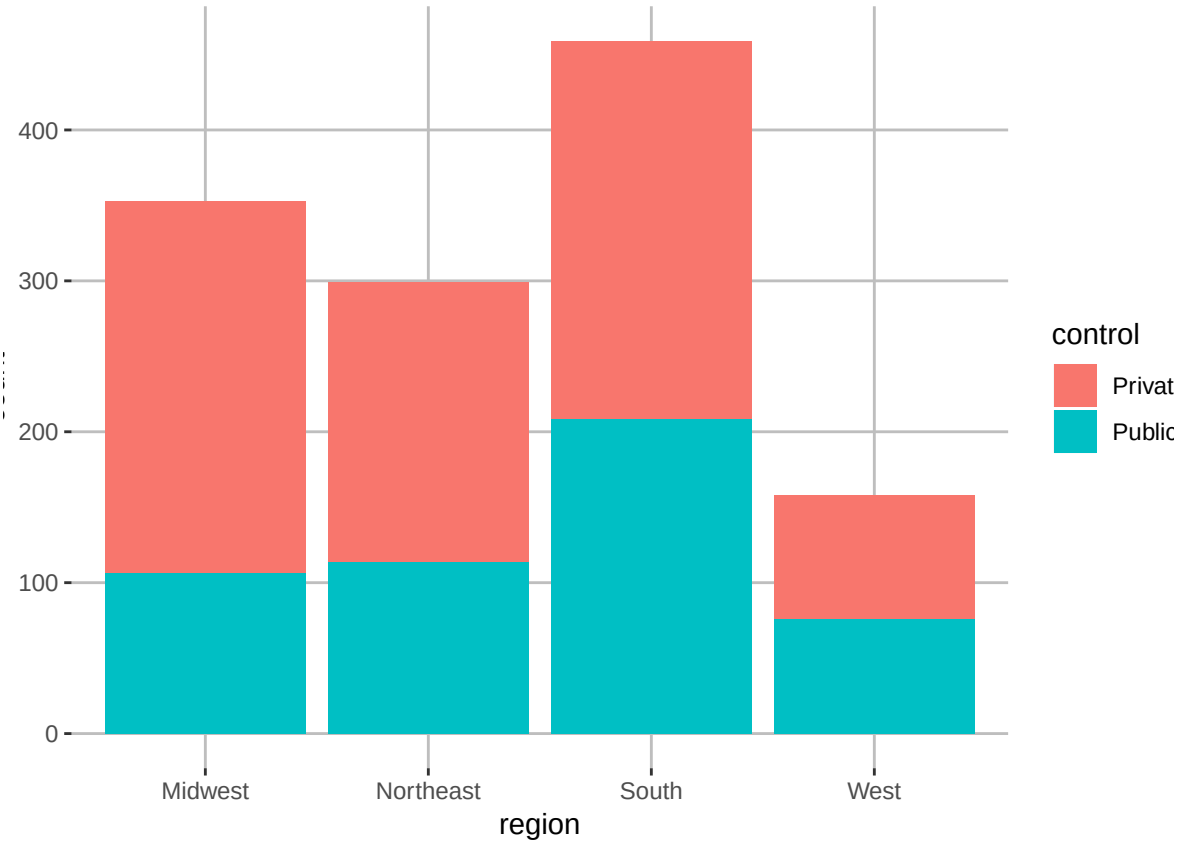
Minimize both backgrounds

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank())
```



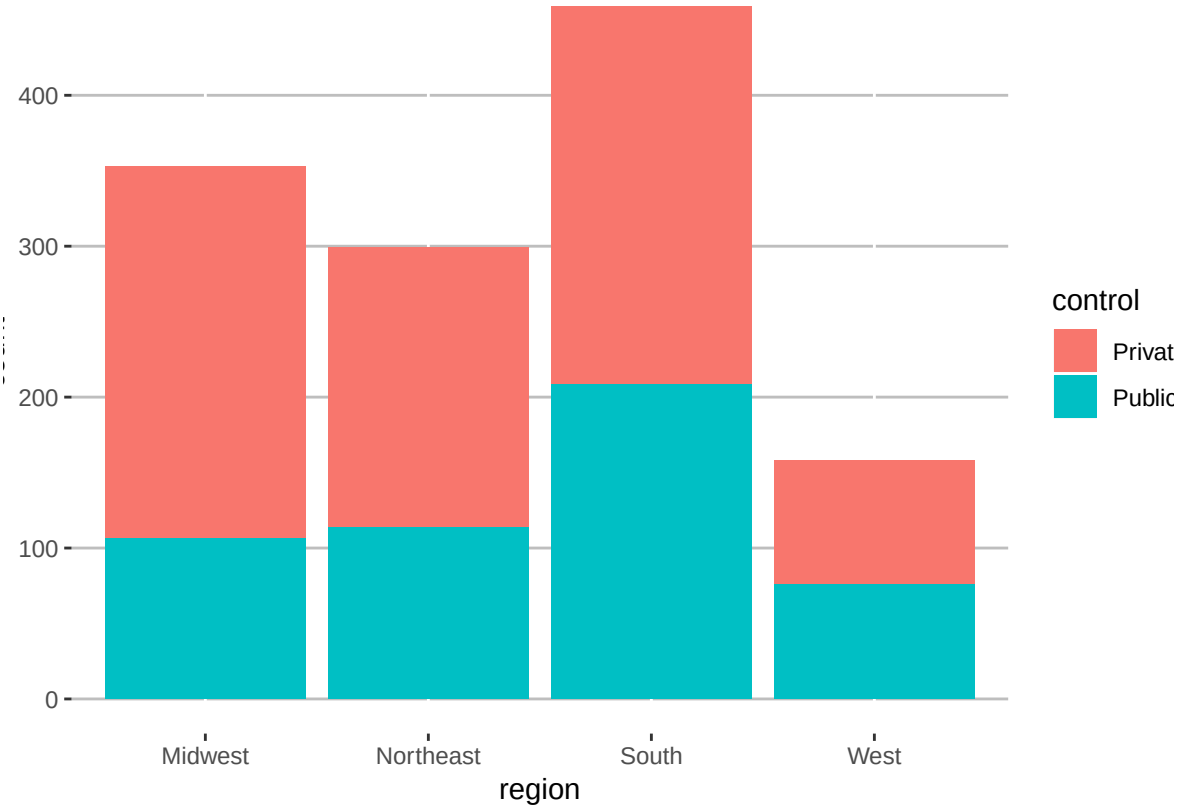
Add grey gridlines to "panel"

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  theme(panel.grid.major=element_line(color="grey"))
```



Show the y-axis gridlines only

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  theme(panel.grid.major.y=element_line(color="grey"))
```



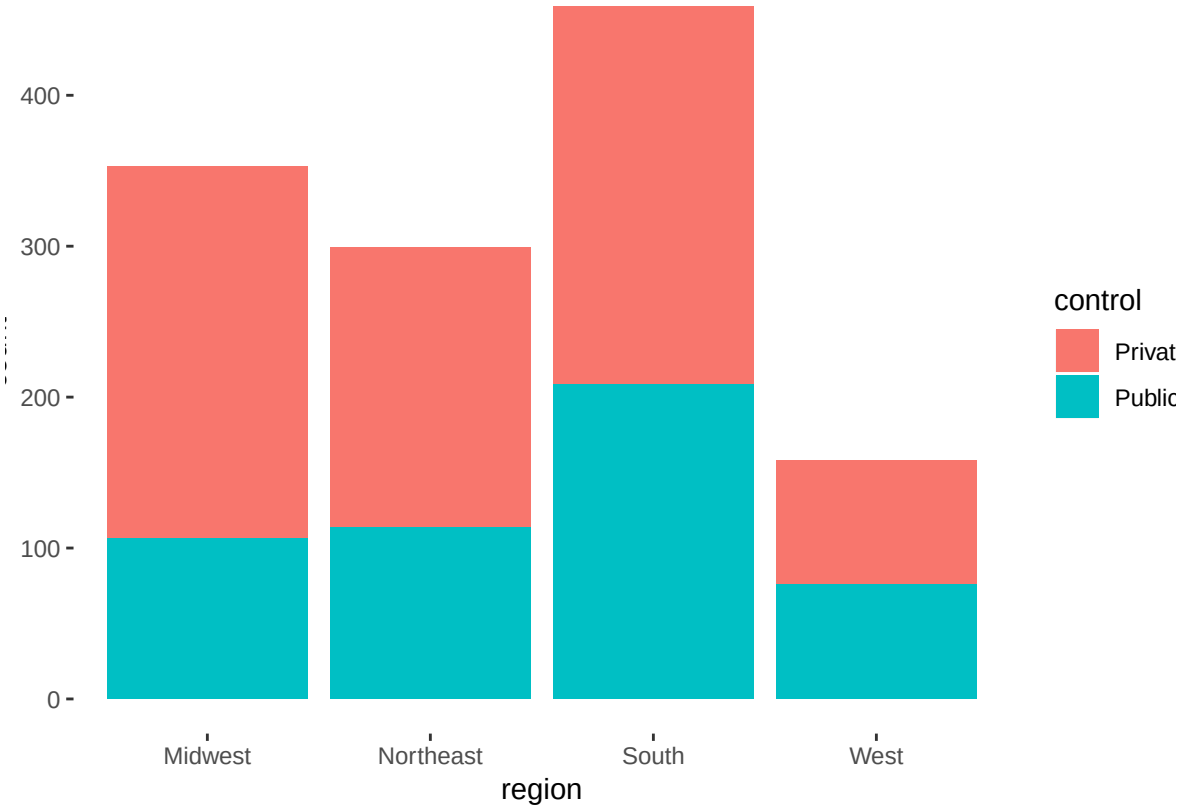
Modifying Axes

Load the dataset

```
library(tidyverse)
college <- read_csv('http://672258.youcanlearnit.net/college.csv')
college <- college %>%
  mutate(state=as.factor(state), region=as.factor(region),
         highest_degree=as.factor(highest_degree),
         control=as.factor(control), gender=as.factor(gender),
         loan_default_rate=as.numeric(loan_default_rate))
```

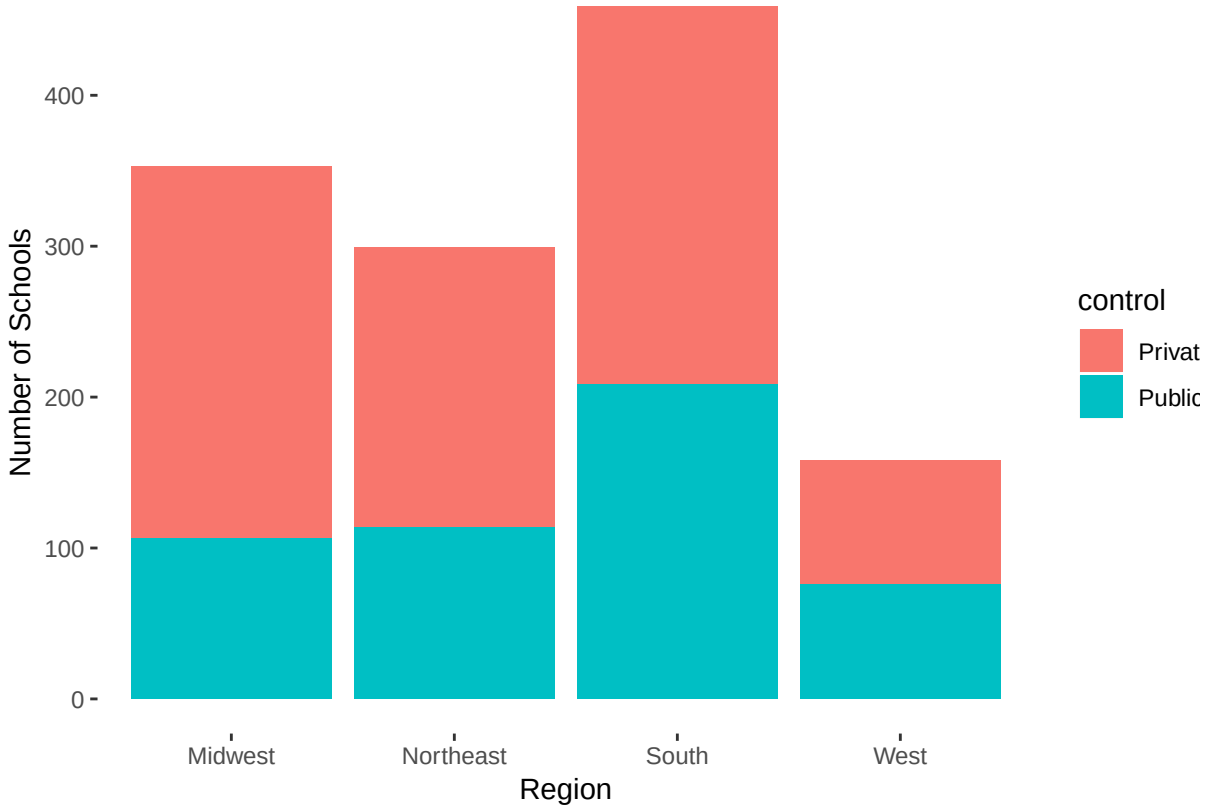
Create the bar graph

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank())
```



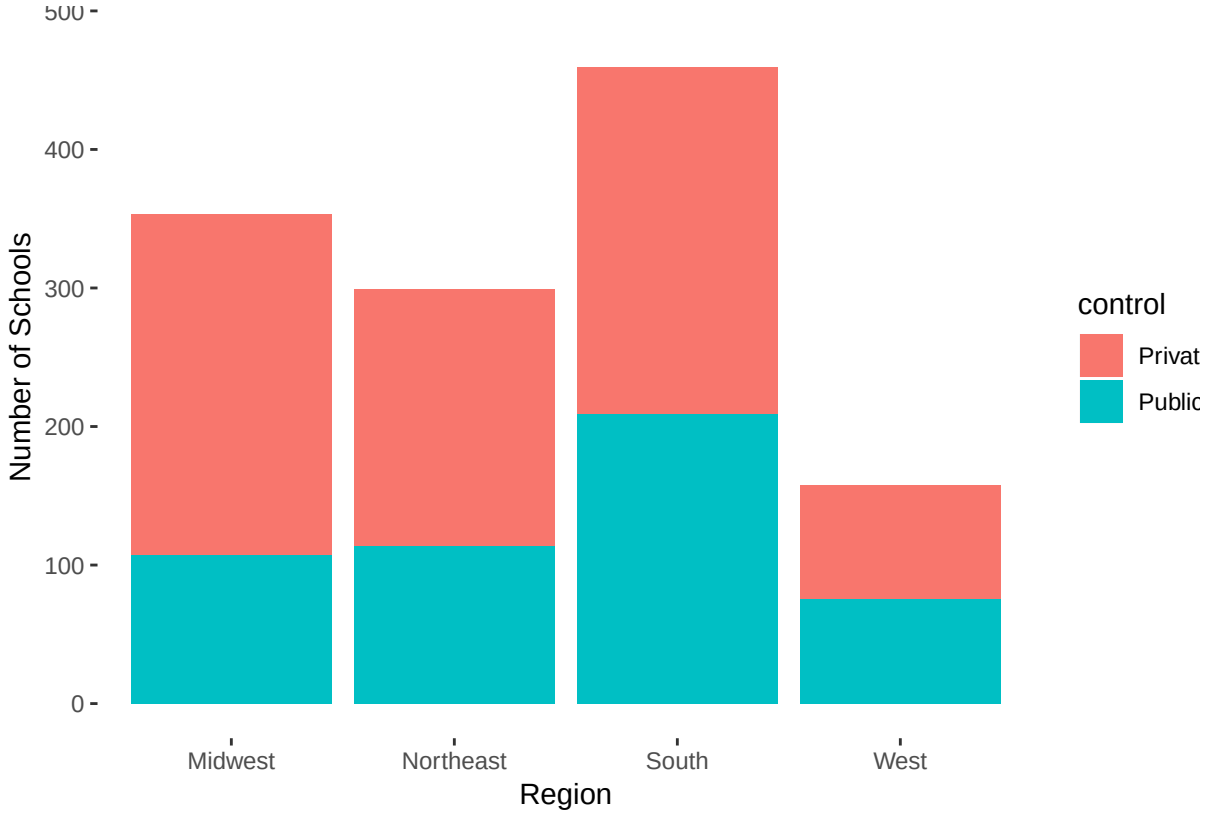
Rename the axes

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  xlab("Region") +
  ylab("Number of Schools")
```



Resize the y-axis

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  xlab("Region") +
  ylab("Number of Schools") +
  ylim(0,500)
```



Modifying Scales

scale_(axis)_(continuous/discrete)

scale_fill_manual

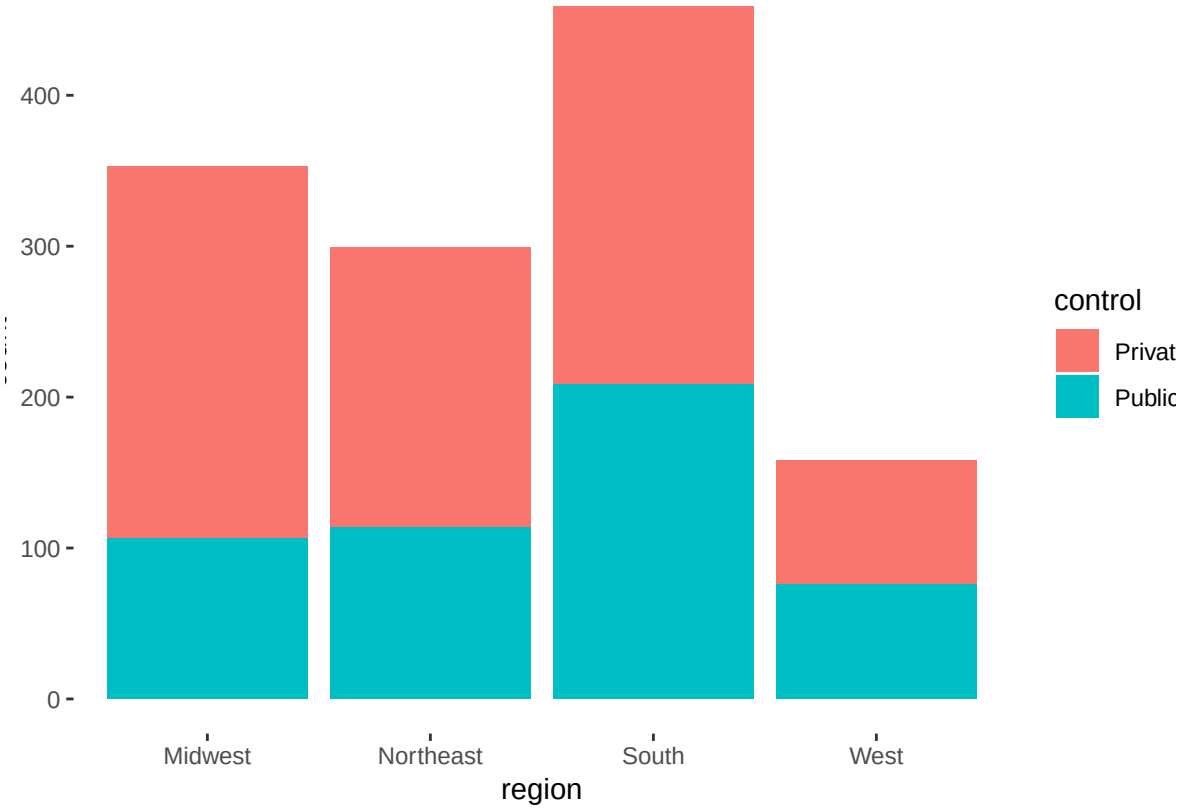
scale_(color/size)_(continuous/discrete)

Load the dataset

```
college <- read_csv('http://672258.youcanlearnit.net/college.csv')
college <- college %>%
  mutate(state=as.factor(state), region=as.factor(region),
         highest_degree=as.factor(highest_degree),
         control=as.factor(control), gender=as.factor(gender),
         loan_default_rate=as.numeric(loan_default_rate))
```

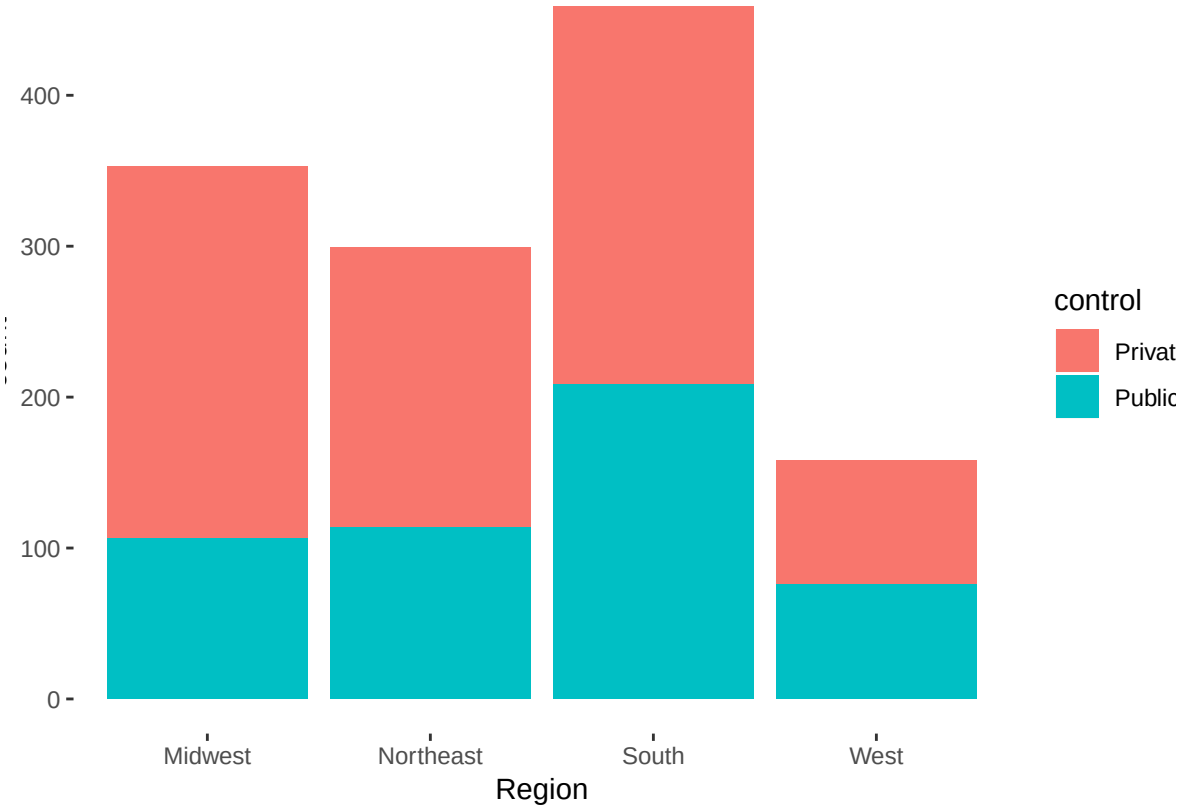
Create the bar graph

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank())
```



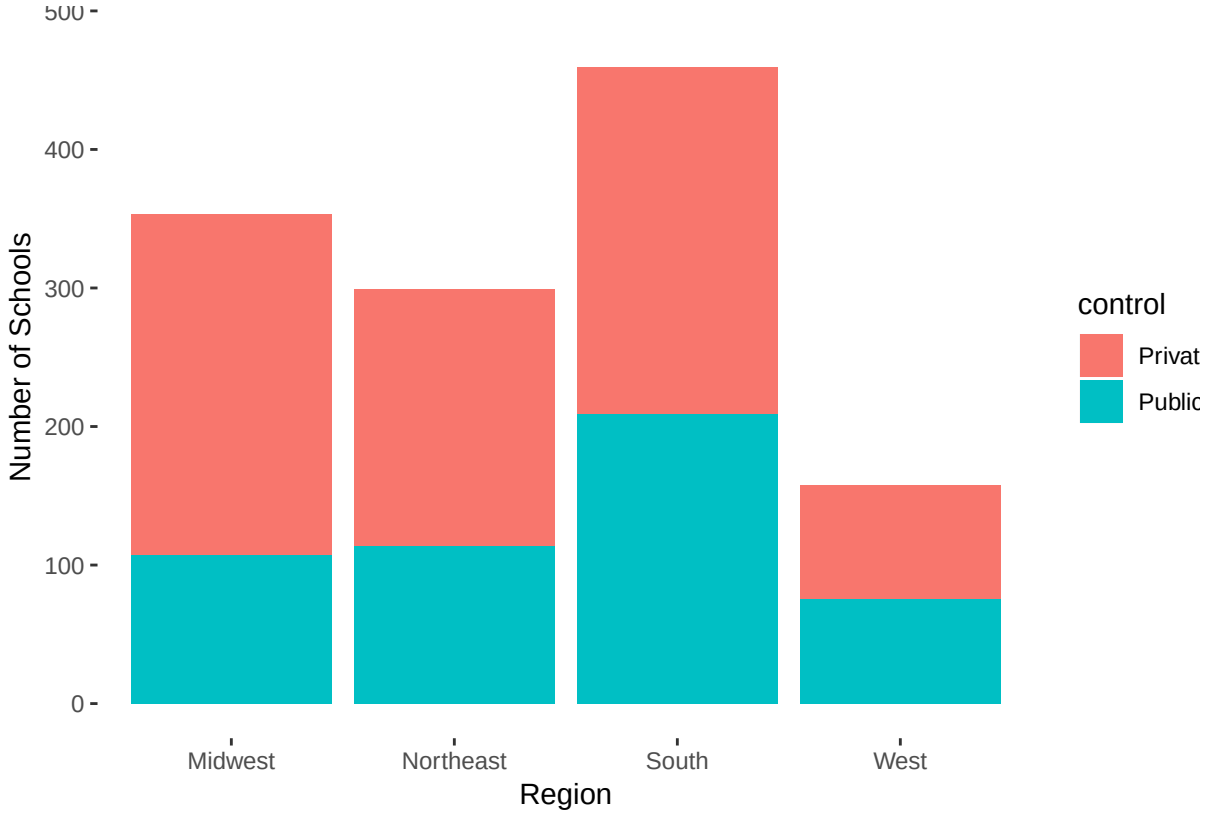
Use scale_ to change the name of x-axis

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region")
```



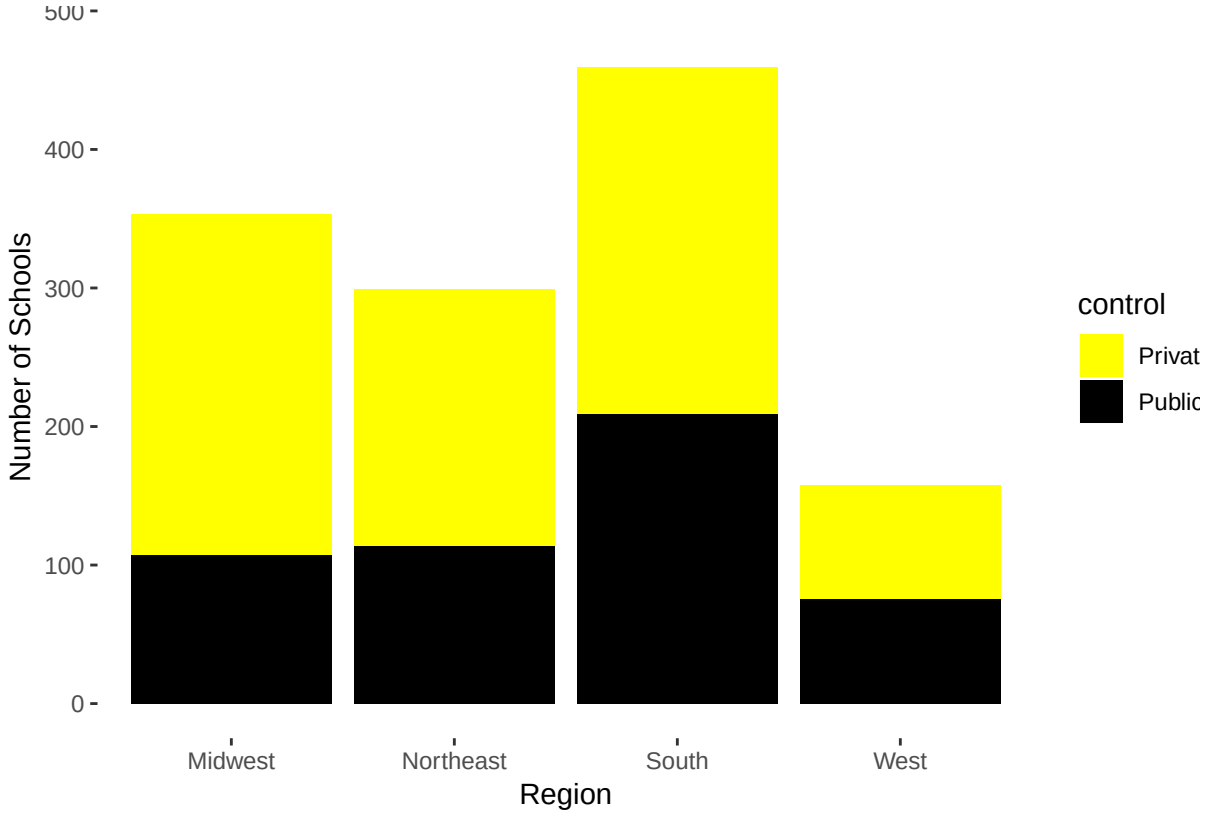
Use scale_ to change the name and limits of the y-axis

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500))
```



Use scale_"fill"_manual to change the fill colors

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"))
```



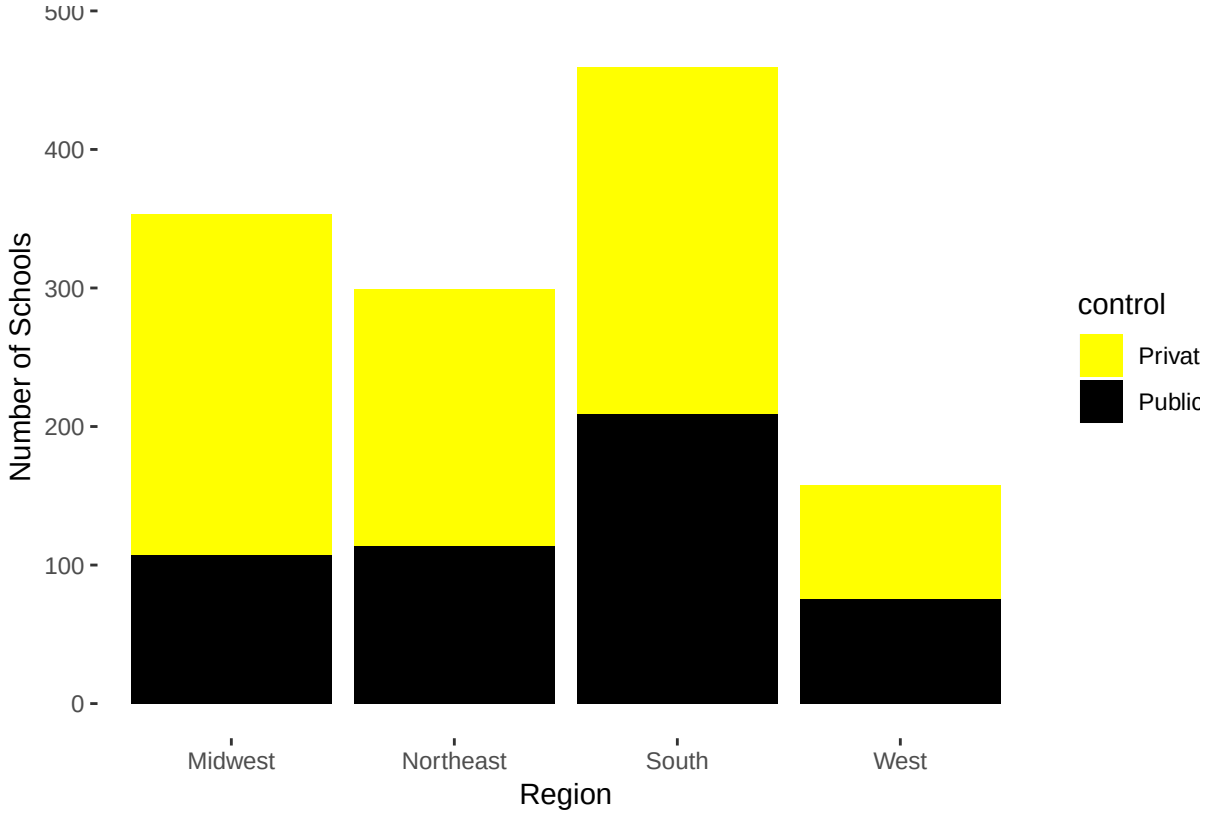
Modifying Legends

Load the dataset

```
college <- read_csv('http://672258.youcanlearnit.net/college.csv')
college <- college %>%
  mutate(state=as.factor(state), region=as.factor(region),
         highest_degree=as.factor(highest_degree),
         control=as.factor(control), gender=as.factor(gender),
         loan_default_rate=as.numeric(loan_default_rate))
```

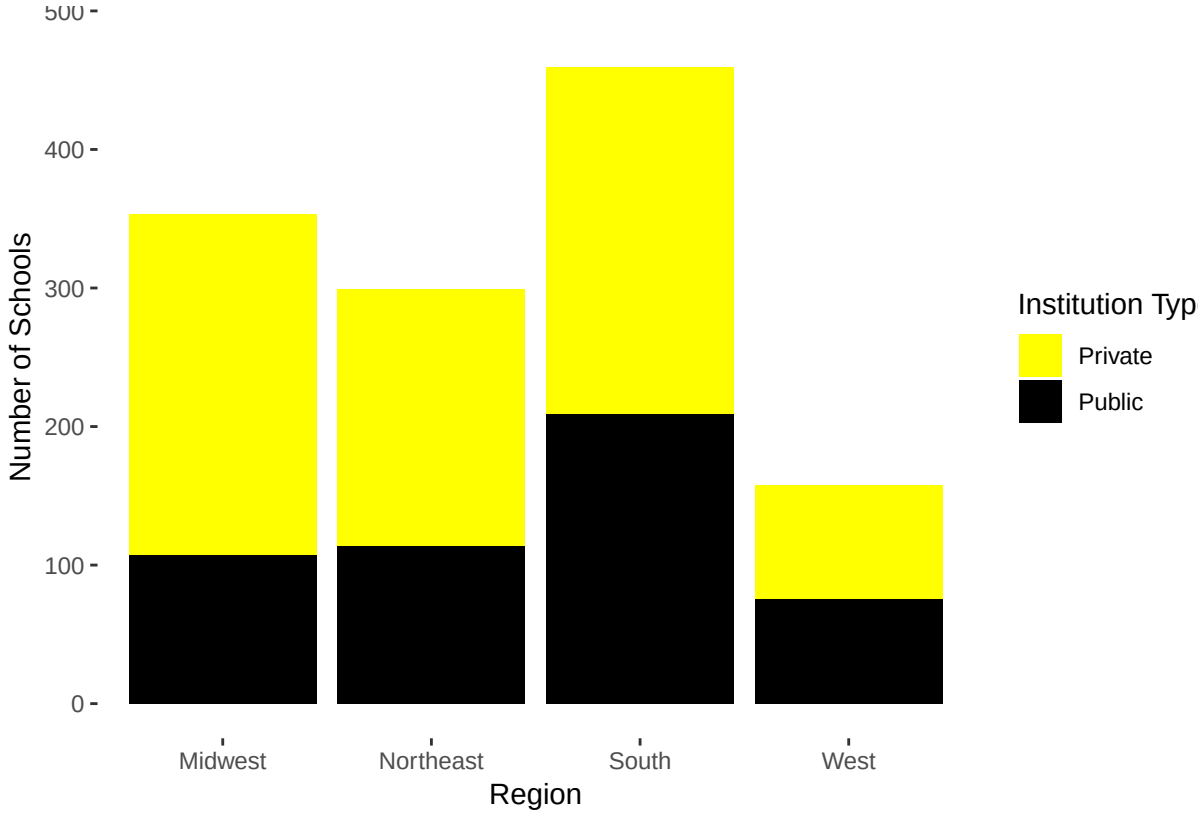
Create the bar graph

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"))
```



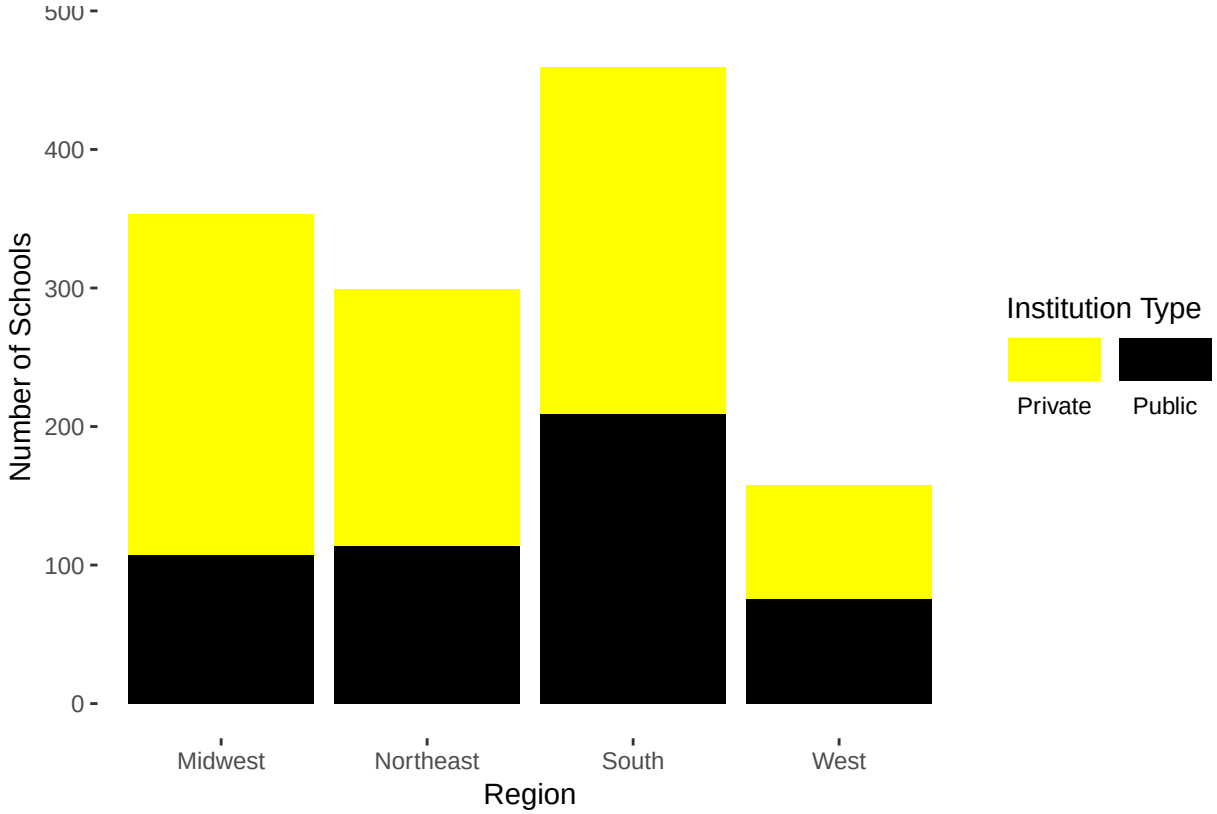
Use scale_"fill"_manual to change the legend title

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"), guide_legend(title="Institution Type"))
```



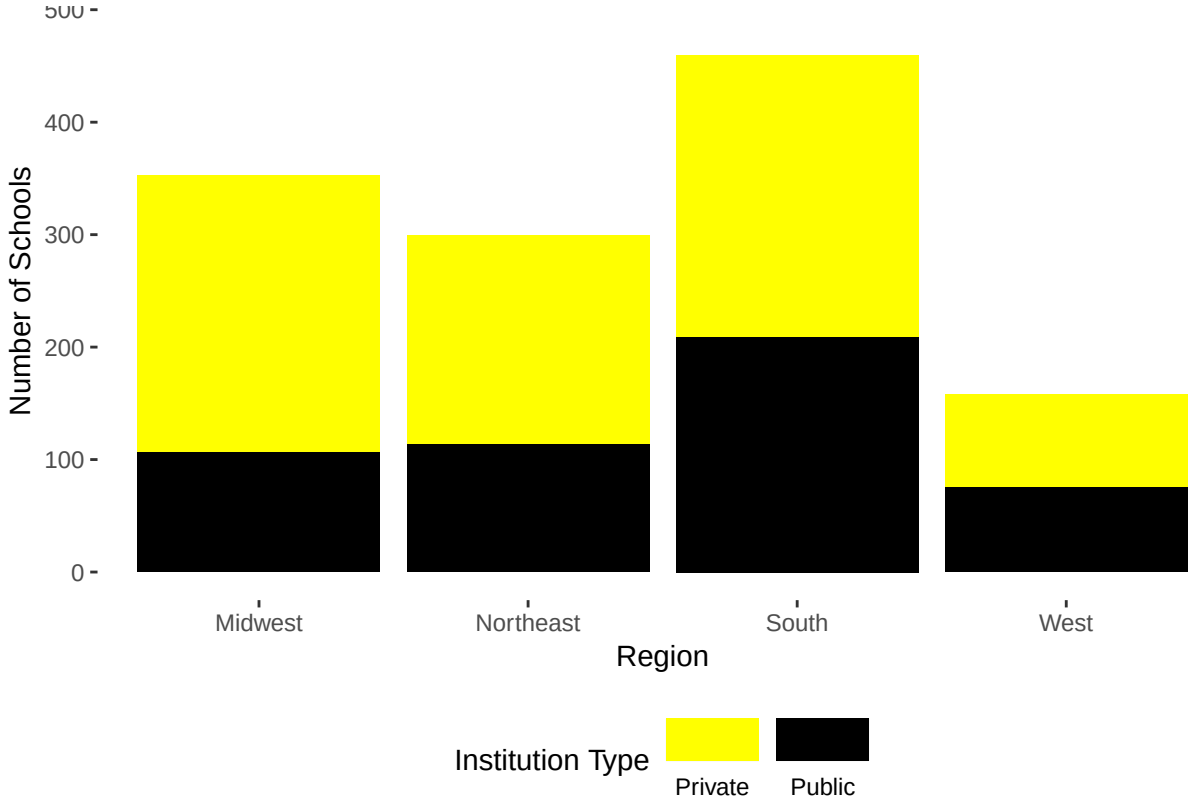
Use scale_"fill"_manual to adjust the legend formatting

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"),
                    guide=guide_legend(title="Institution Type", label.position="bottom", nrow=1, keywi
```



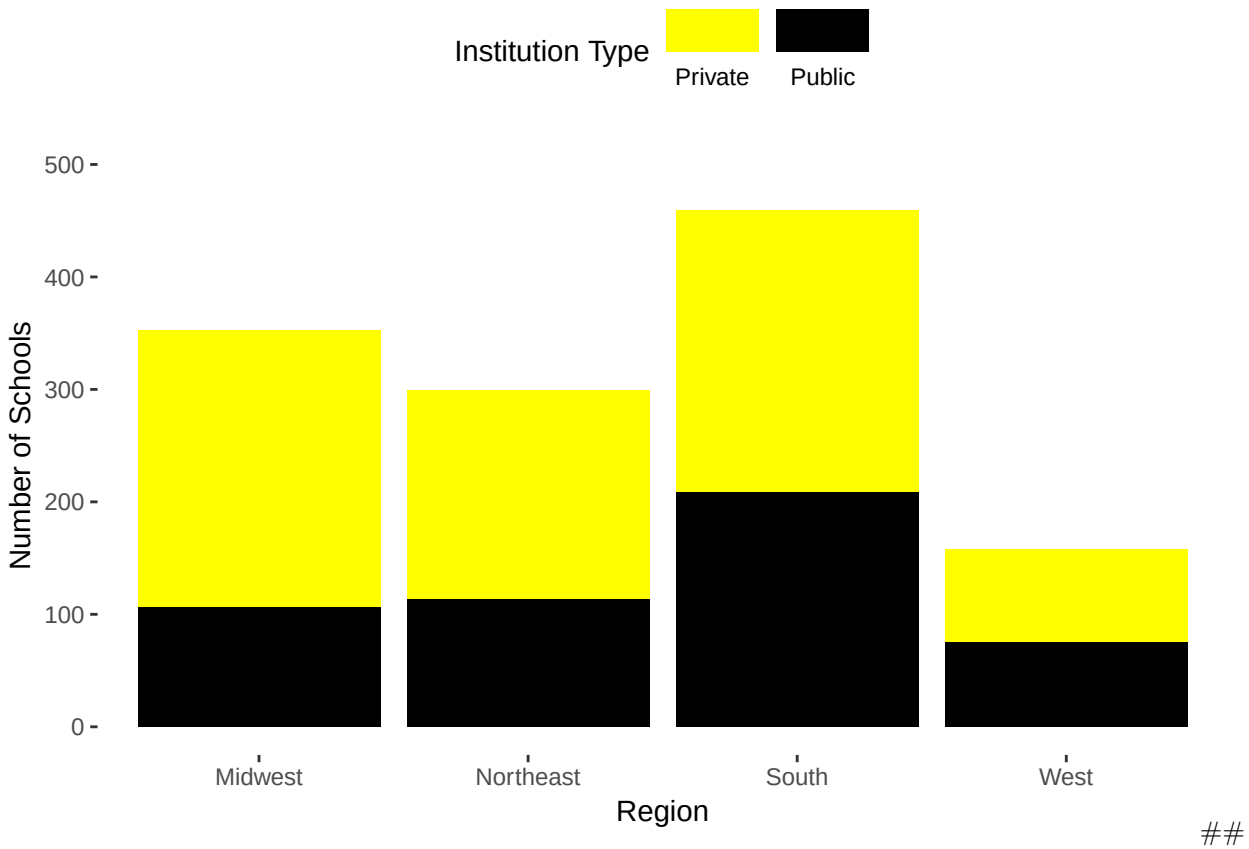
Use "theme" to move the legend to the bottom of the plot

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"),
                    guide=guide_legend(title="Institution Type", label.position="bottom", nrow=1, keywi
  theme(legend.position="bottom")
```



Use "theme" to move the legend to the top of the plot

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"),
                    guide=guide_legend(title="Institution Type", label.position="bottom", nrow=1, keywi
  theme(legend.position="top")
```



##

Annotating Plots

Load the dataset

```
college <- read_csv('http://672258.youcanlearnit.net/college.csv')
college <- college %>%
  mutate(state=as.factor(state), region=as.factor(region),
         highest_degree=as.factor(highest_degree),
         control=as.factor(control), gender=as.factor(gender),
         loan_default_rate=as.numeric(loan_default_rate))
```

Create the scatterplot

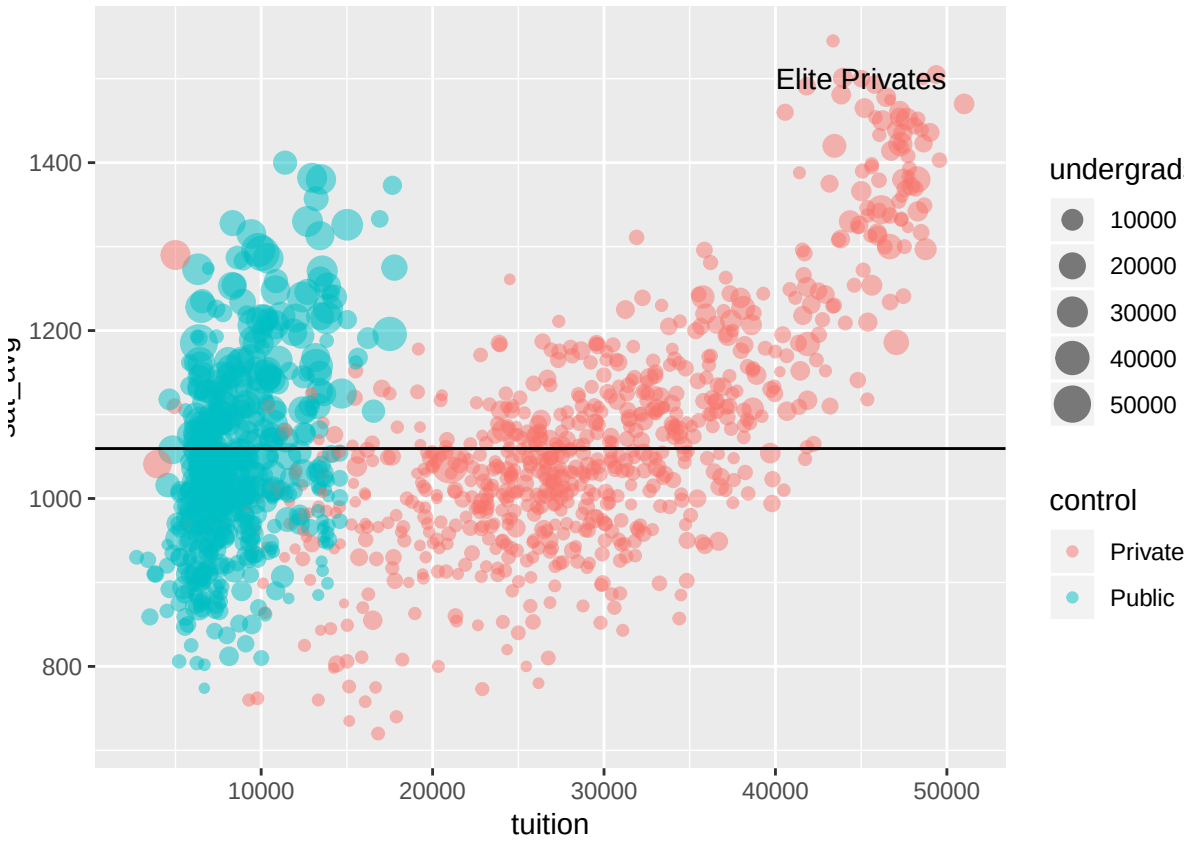
```
ggplot(data=college) +
  geom_point(mapping=aes(x=tuition, y=sat_avg, color=control, size=undergrads), alpha=0.5)
```

Use annotate() to add a text annotation

```
ggplot(data=college) +
  geom_point(mapping=aes(x=tuition, y=sat_avg, color=control, size=undergrads), alpha=0.5) +
  annotate("text", label="Elite Privates", x=45000,y=1500)
```

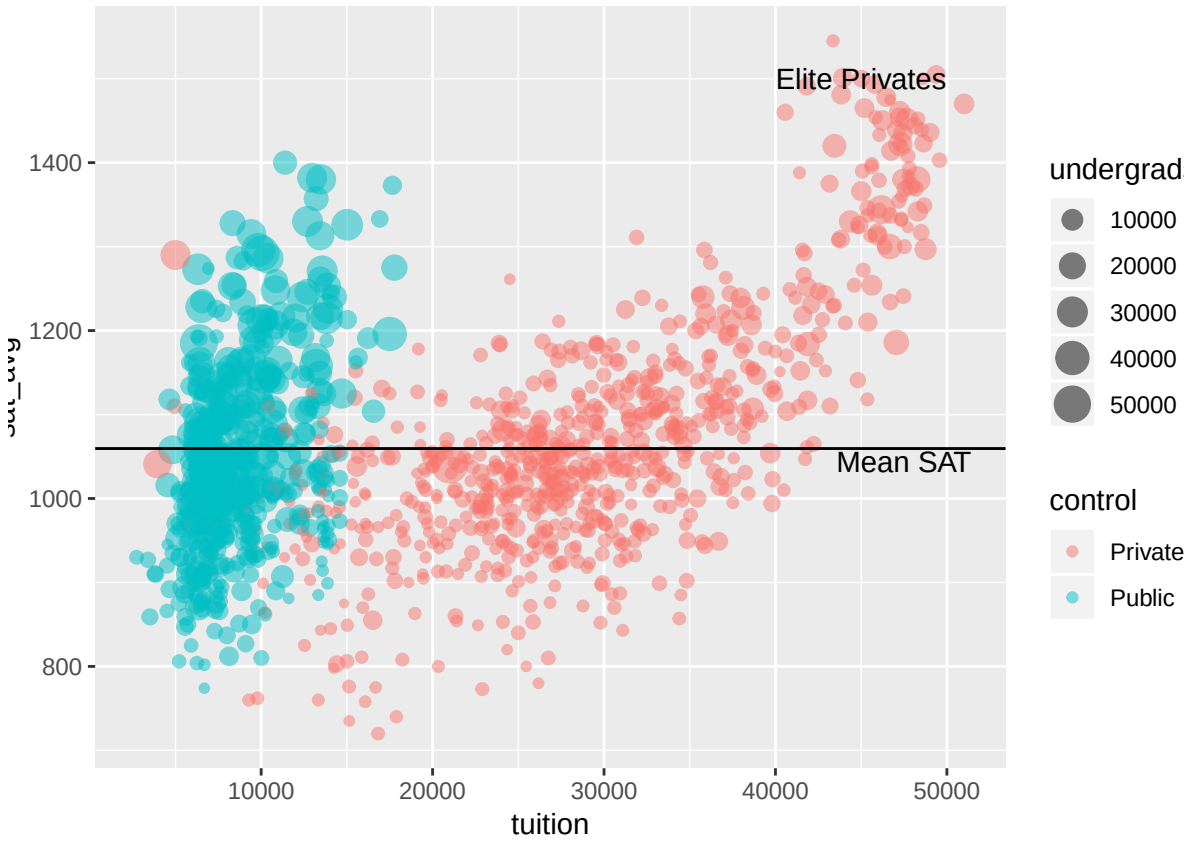
Use geom_hline() to add a line for the mean SAT score

```
ggplot(data=college) +
  geom_point(mapping=aes(x=tuition, y=sat_avg, color=control, size=undergrads), alpha=0.5) +
  annotate("text", label="Elite Privates", x=45000,y=1500) +
  geom_hline(yintercept=mean(college$sat_avg))
```



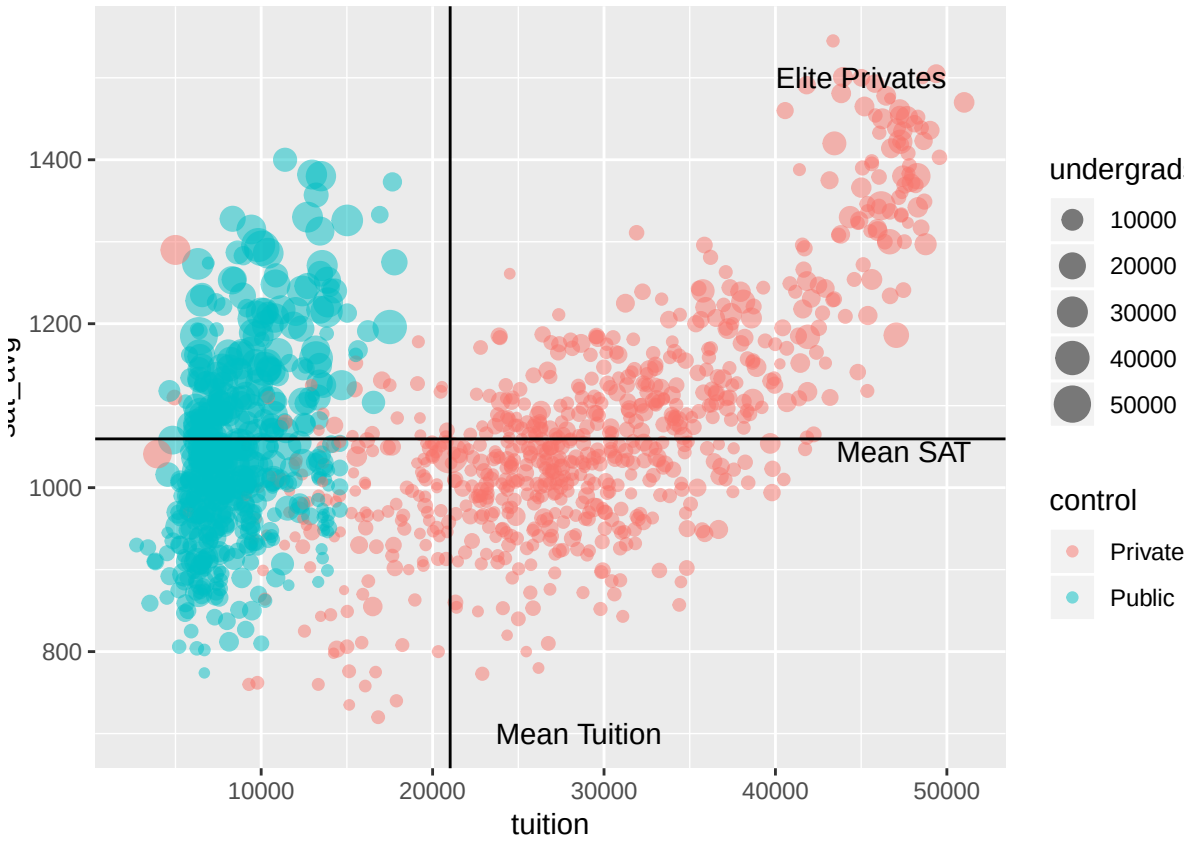
Use geom_hline() to label the line

```
ggplot(data=college) +
  geom_point(mapping=aes(x=tuition, y=sat_avg, color=control, size=undergrads), alpha=0.5) +
  annotate("text", label="Elite Privates", x=45000,y=1500) +
  geom_hline(yintercept=mean(college$sat_avg)) +
  annotate("text", label="Mean SAT", x=47500, y=mean(college$sat_avg)-15)
```



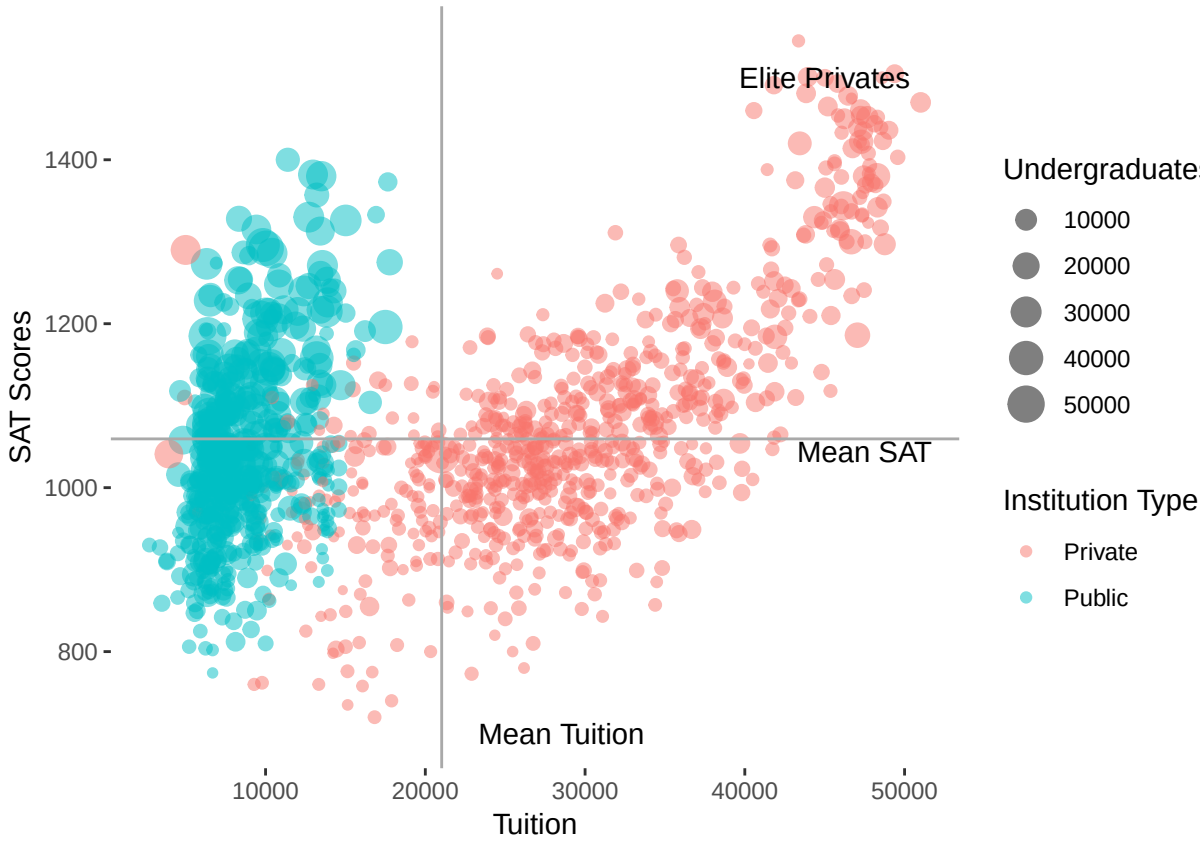
Use geom_Vline() to add a line for mean tuition and add a text annotation

```
ggplot(data=college) +
  geom_point(mapping=aes(x=tuition, y=sat_avg, color=control, size=undergrads), alpha=0.5) +
  annotate("text", label="Elite Privates", x=45000,y=1500) +
  geom_hline(yintercept=mean(college$sat_avg)) +
  annotate("text", label="Mean SAT", x=47500, y=mean(college$sat_avg)-15) +
  geom_vline(xintercept=mean(college$tuition)) +
  annotate("text", label="Mean Tuition", x=mean(college$tuition)+7500, y=700)
```



Use scale_to change names

```
ggplot(data=college) +
  geom_point(mapping=aes(x=tuition, y=sat_avg, color=control, size=undergrads), alpha=0.5) +
  annotate("text", label="Elite Privates", x=45000,y=1500) +
  geom_hline(yintercept=mean(college$sat_avg), color="dark grey") +
  annotate("text", label="Mean SAT", x=47500, y=mean(college$sat_avg)-15) +
  geom_vline(xintercept=mean(college$tuition), color="dark grey") +
  annotate("text", label="Mean Tuition", x=mean(college$tuition)+7500, y=700) +
  theme(panel.background = element_blank(), legend.key = element_blank()) +
  scale_color_discrete(name="Institution Type") +
  scale_size_continuous(name="Undergraduates") +
  scale_x_continuous(name="Tuition") +
  scale_y_continuous(name="SAT Scores")
```



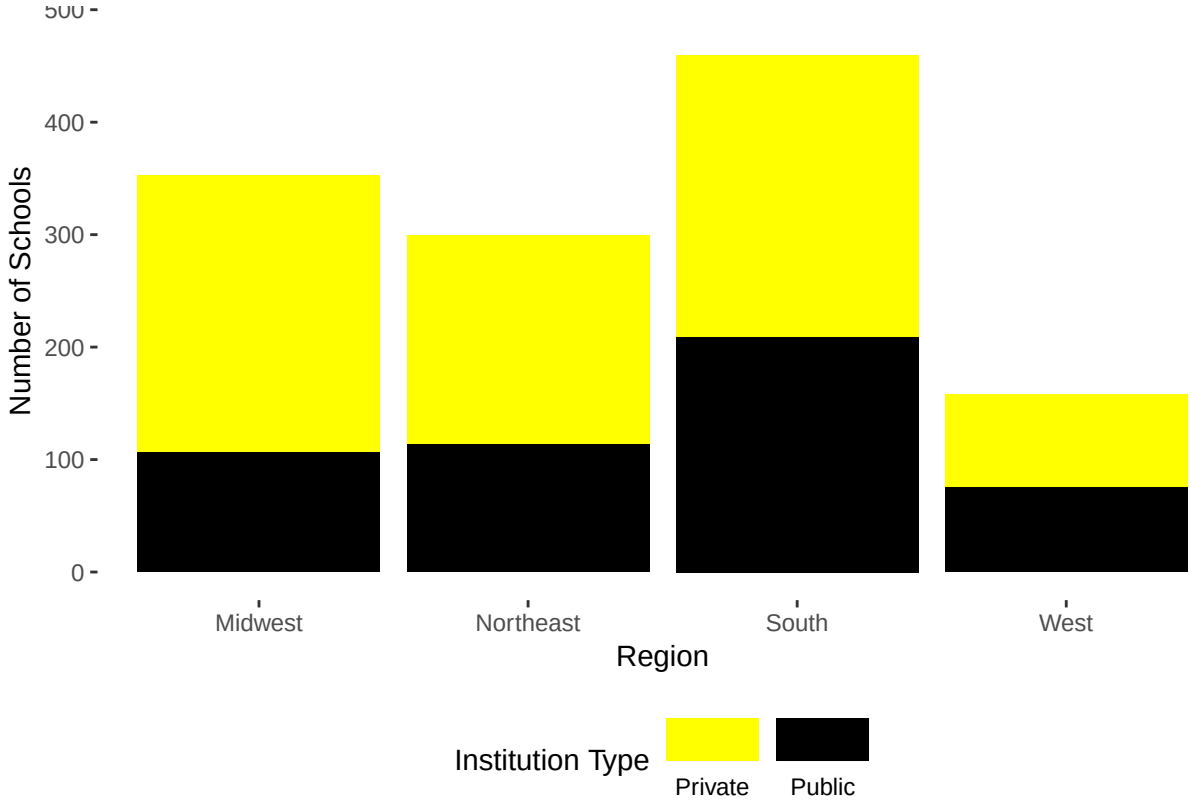
Titles

Load the dataset

```
library(tidyverse)
college <- read_csv('http://672258.youcanlearnit.net/college.csv')
college <- college %>%
  mutate(state=as.factor(state), region=as.factor(region),
         highest_degree=as.factor(highest_degree),
         control=as.factor(control), gender=as.factor(gender),
         loan_default_rate=as.numeric(loan_default_rate))
```

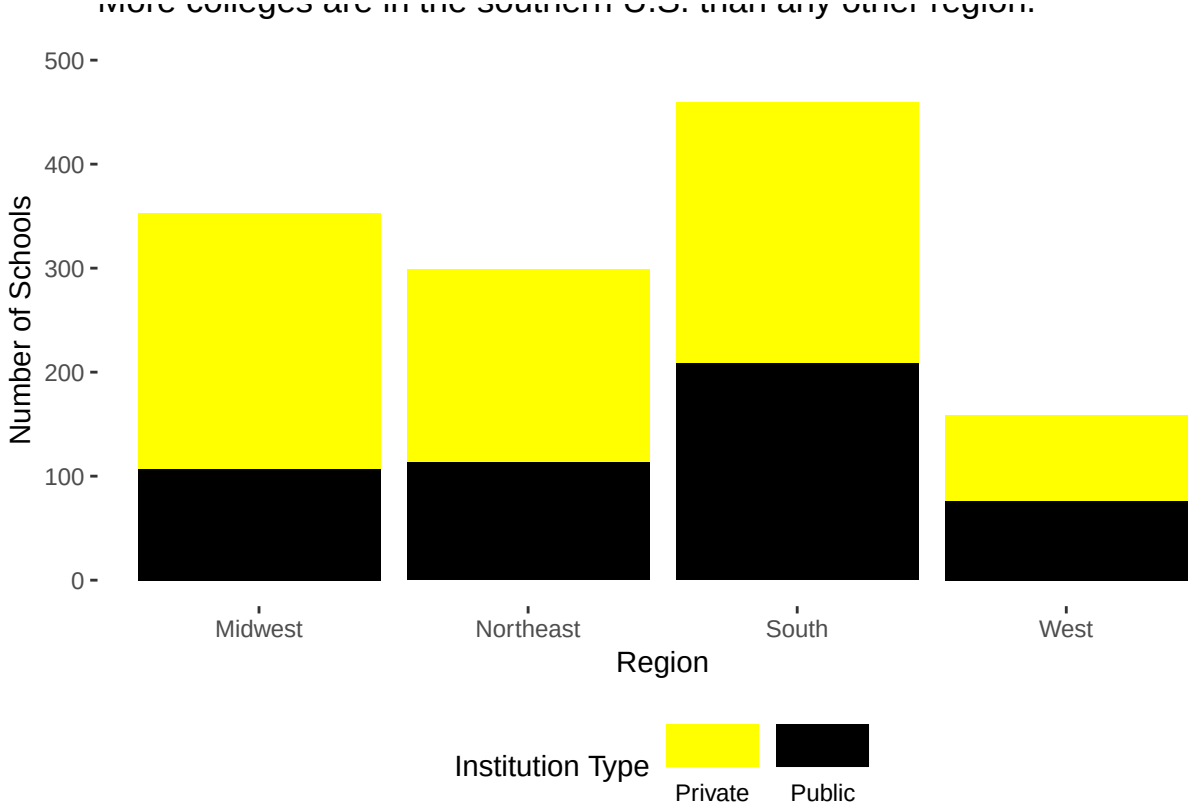
Create the bar graph

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"),
                    guide=guide_legend(title="Institution Type", label.position="bottom", nrow=1, keywi
  theme(legend.position="bottom")
```



Add a title

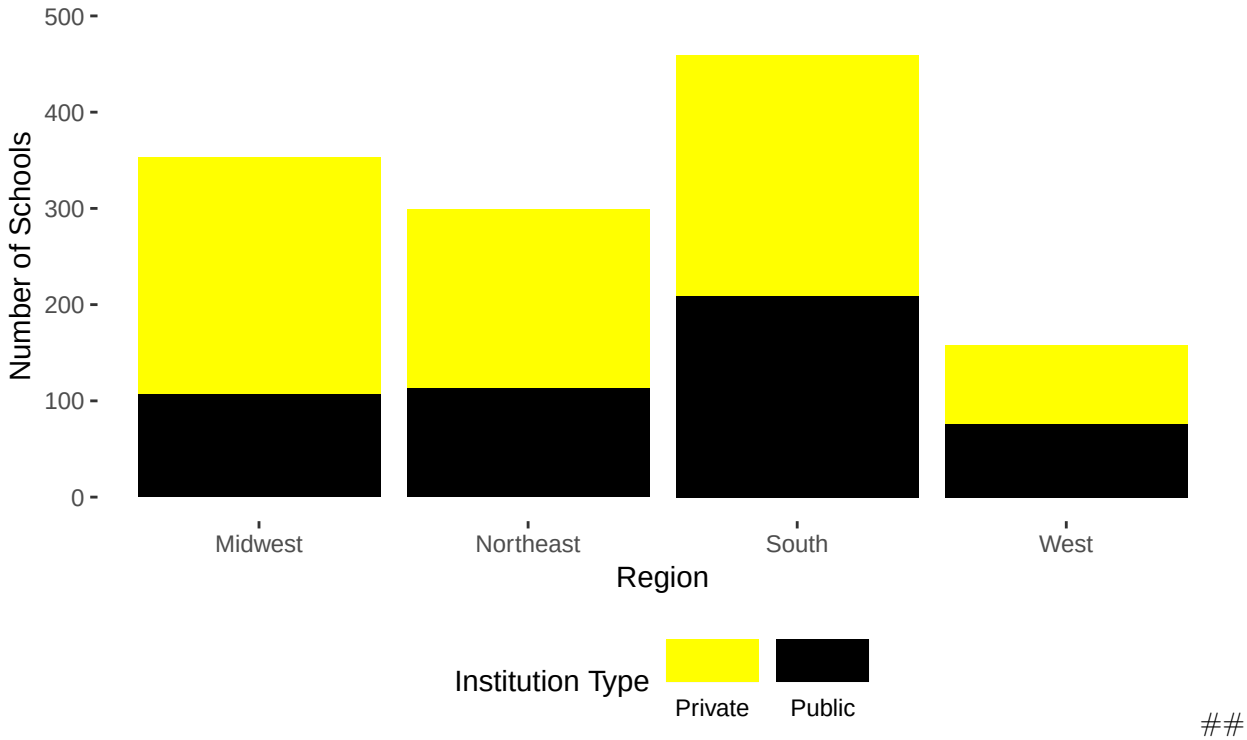
```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"),
                    guide=guide_legend(title="Institution Type", label.position="bottom", nrow=1, keywi
  theme(legend.position="bottom") +
  ggtitle("More colleges are in the southern U.S. than any other region.")
```



Add a subtitle

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme(panel.background=element_blank()) +
  theme(plot.background=element_blank()) +
  scale_x_discrete(name="Region") +
  scale_y_continuous(name="Number of Schools", limits=c(0,500)) +
  scale_fill_manual(values=c("yellow","black"),
                    guide=guide_legend(title="Institution Type", label.position="bottom", nrow=1, keywi
  theme(legend.position="bottom") +
  ggtitle("More colleges are in the southern U.S. than any other region.",
          subtitle="Source: U.S. Department of Education")
```

More colleges are in the southern U.S. than any other region.
Source: U.S. Department of Education



##

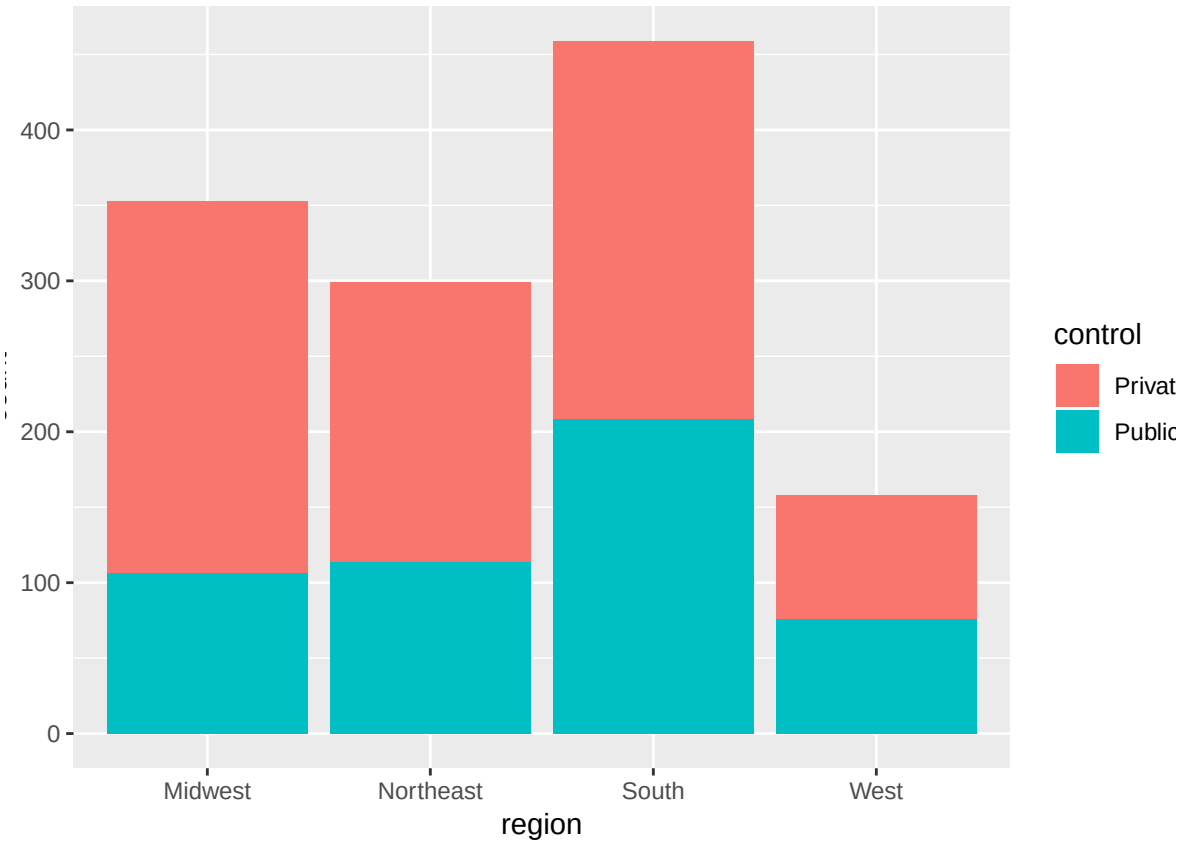
Themes

Load the dataset

```
college <- read_csv('http://672258.youcanlearnit.net/college.csv')
college <- college %>%
  mutate(state=as.factor(state), region=as.factor(region),
         highest_degree=as.factor(highest_degree),
         control=as.factor(control), gender=as.factor(gender),
         loan_default_rate=as.numeric(loan_default_rate))
```

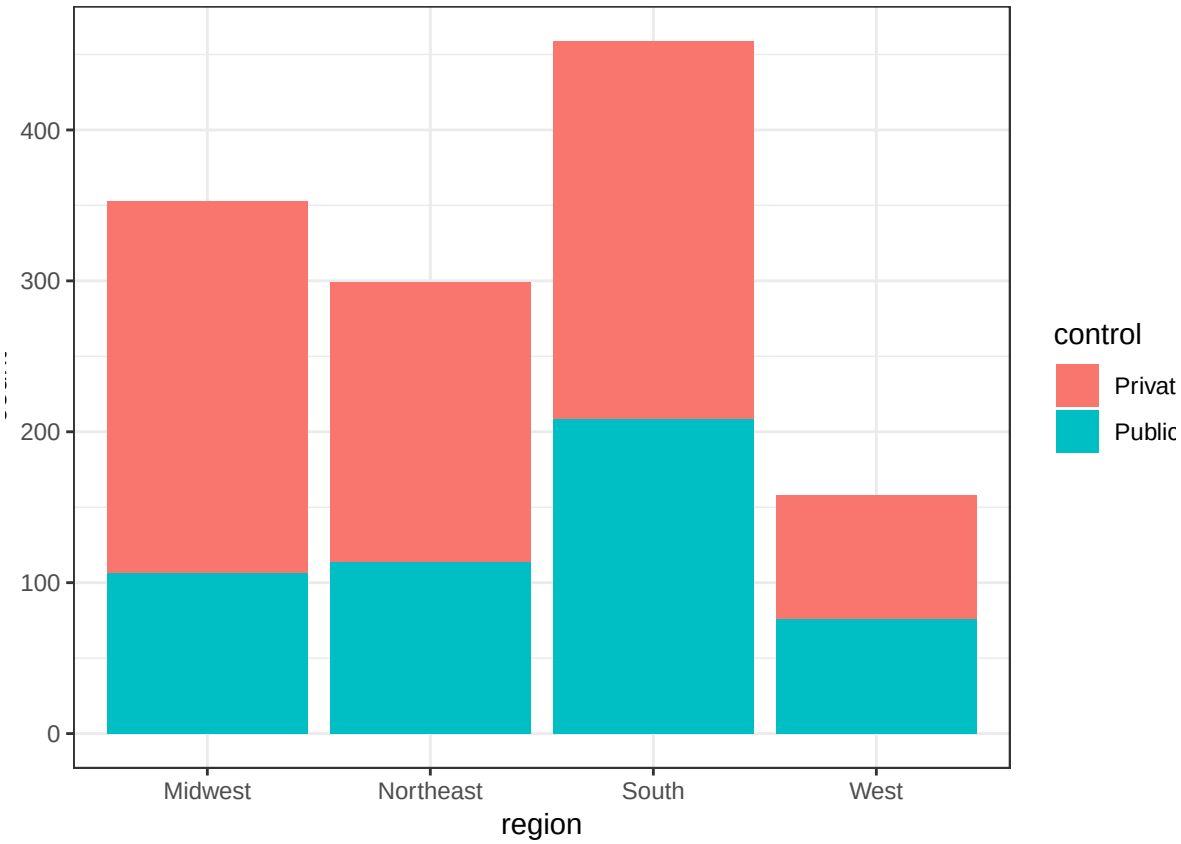
Create the bar graph

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control))
```



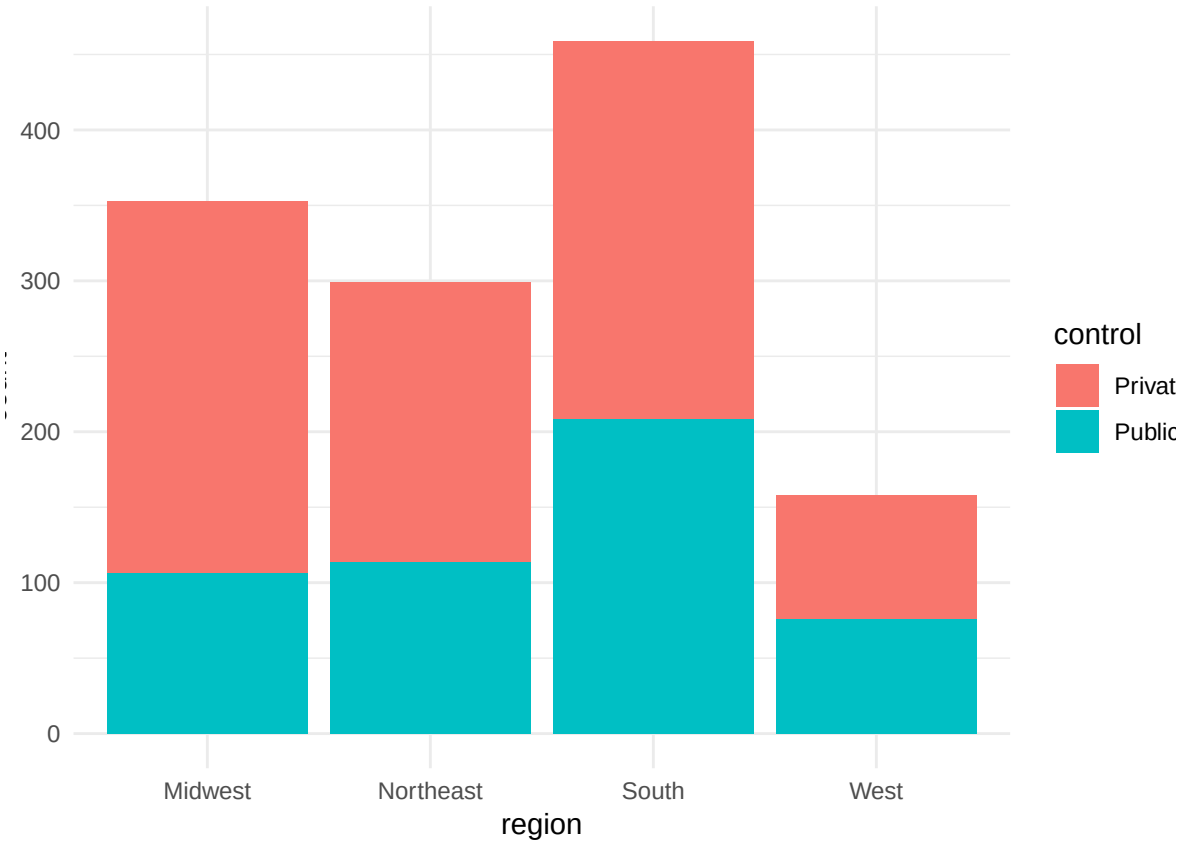
Black and white theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_bw()
```



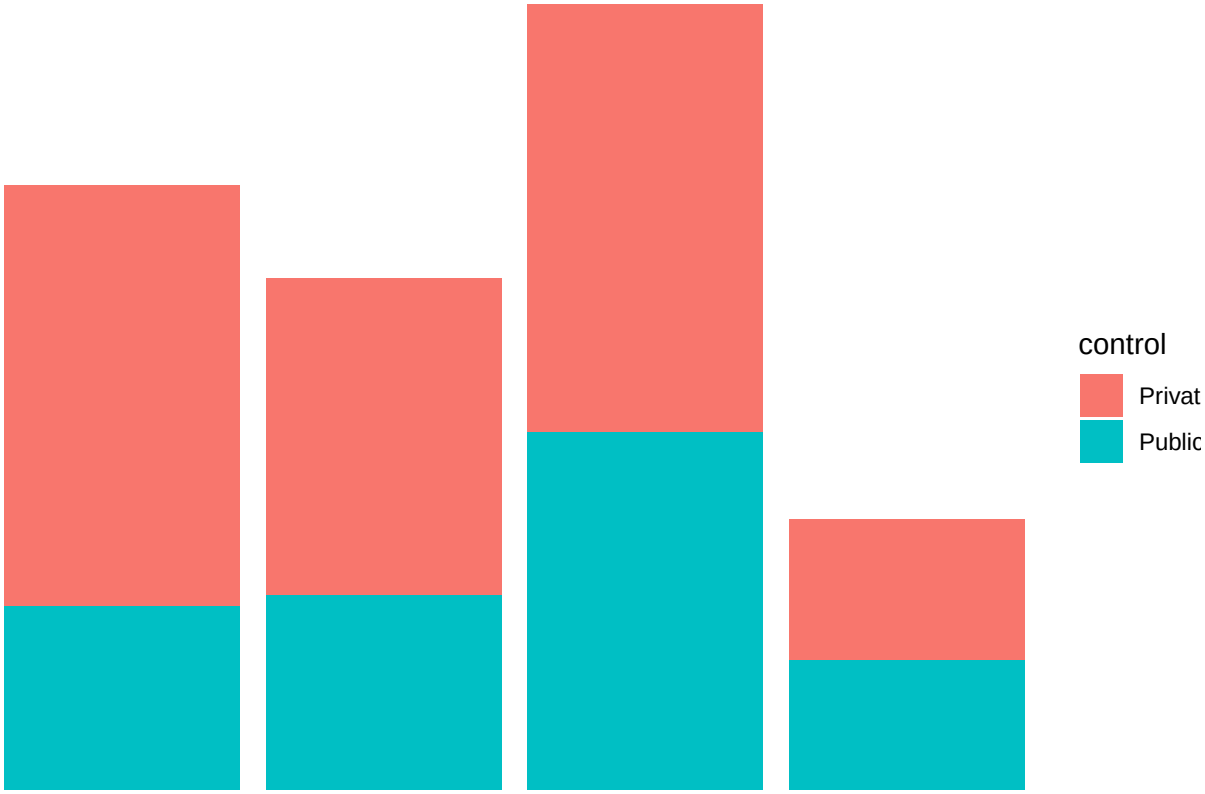
Minimal theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_minimal()
```



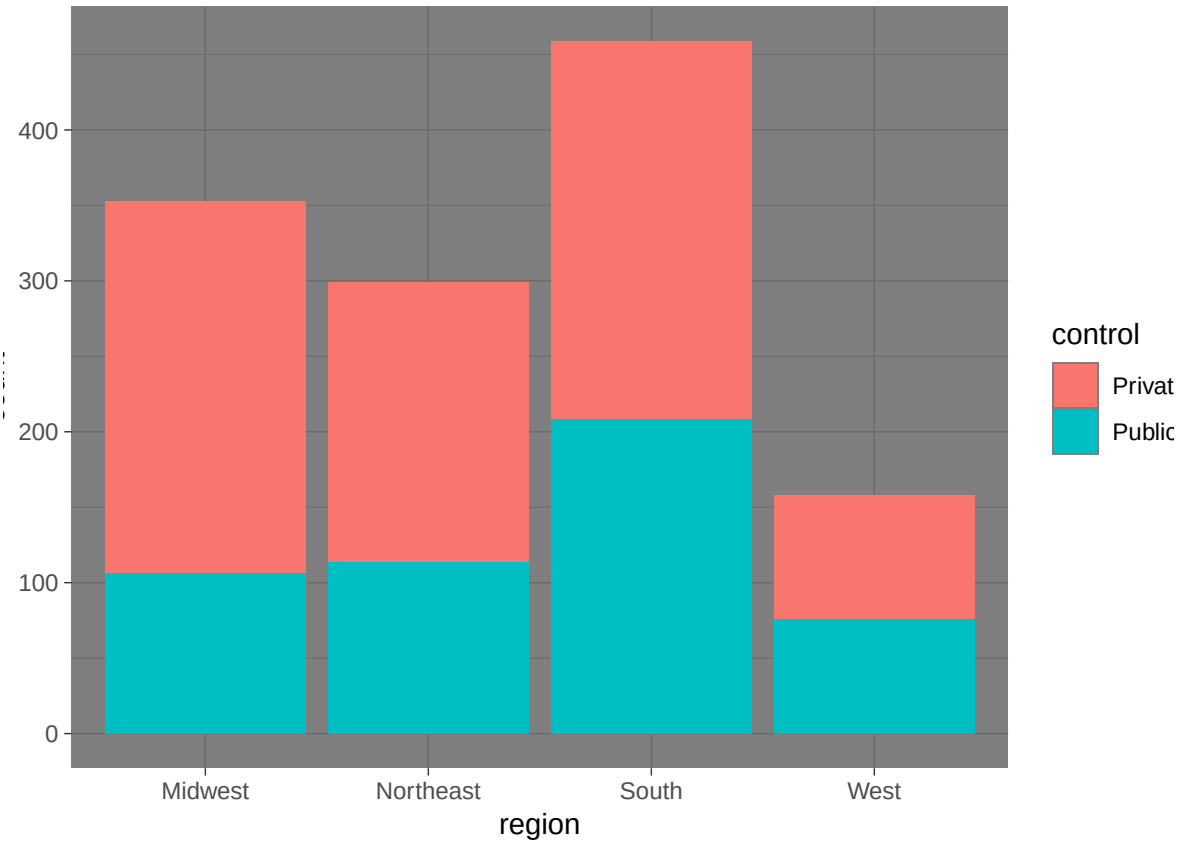
Void theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_void()
```



Dark theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_dark()
```

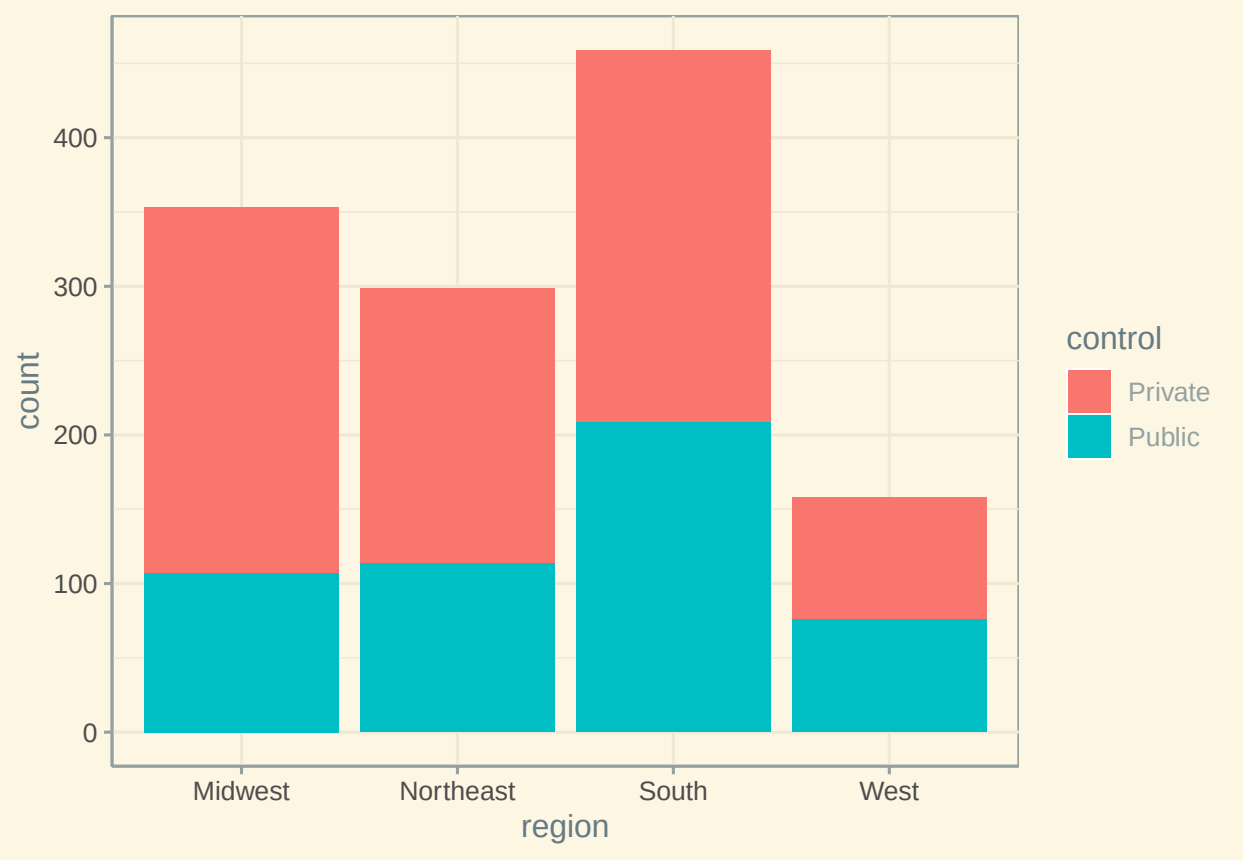


ggthemes Package

```
library(ggthemes)
```

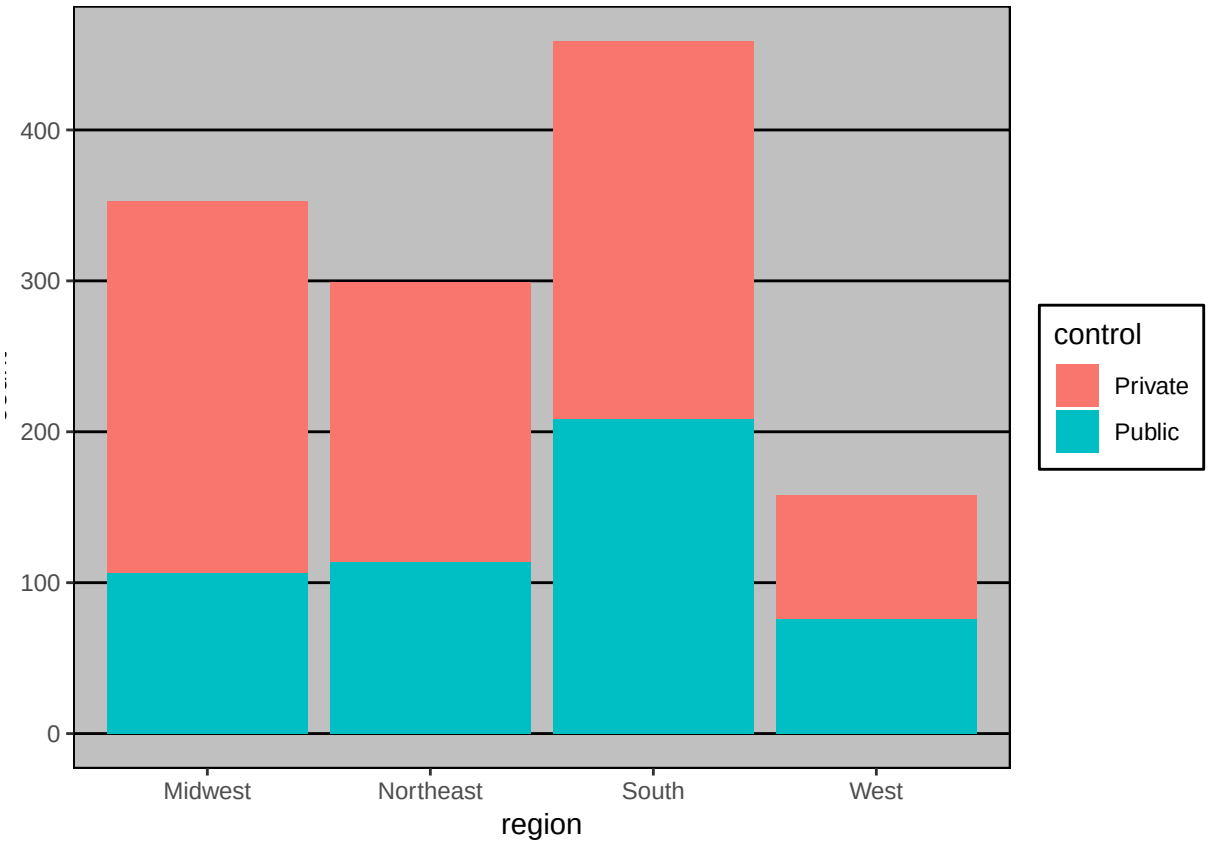
Solarized theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_solarized()
```



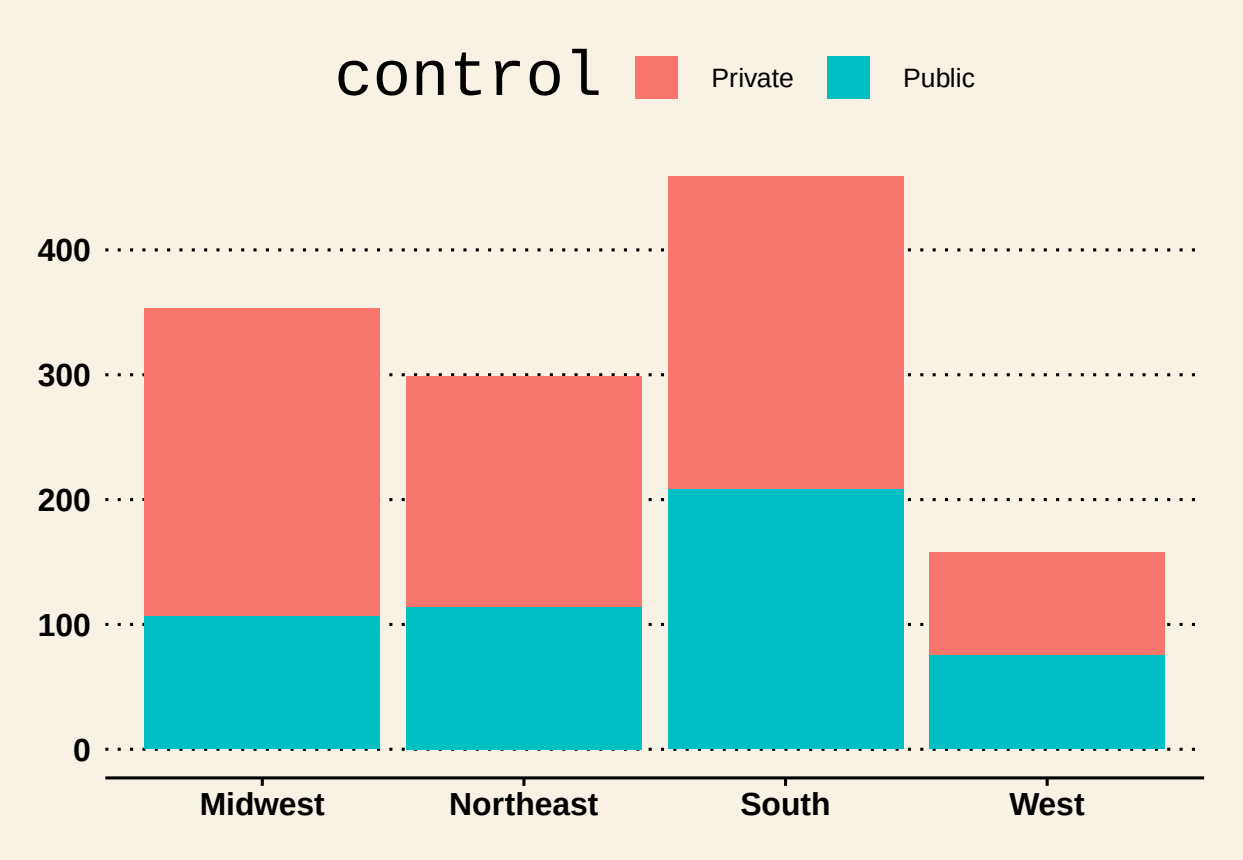
Excel theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_excel()
```



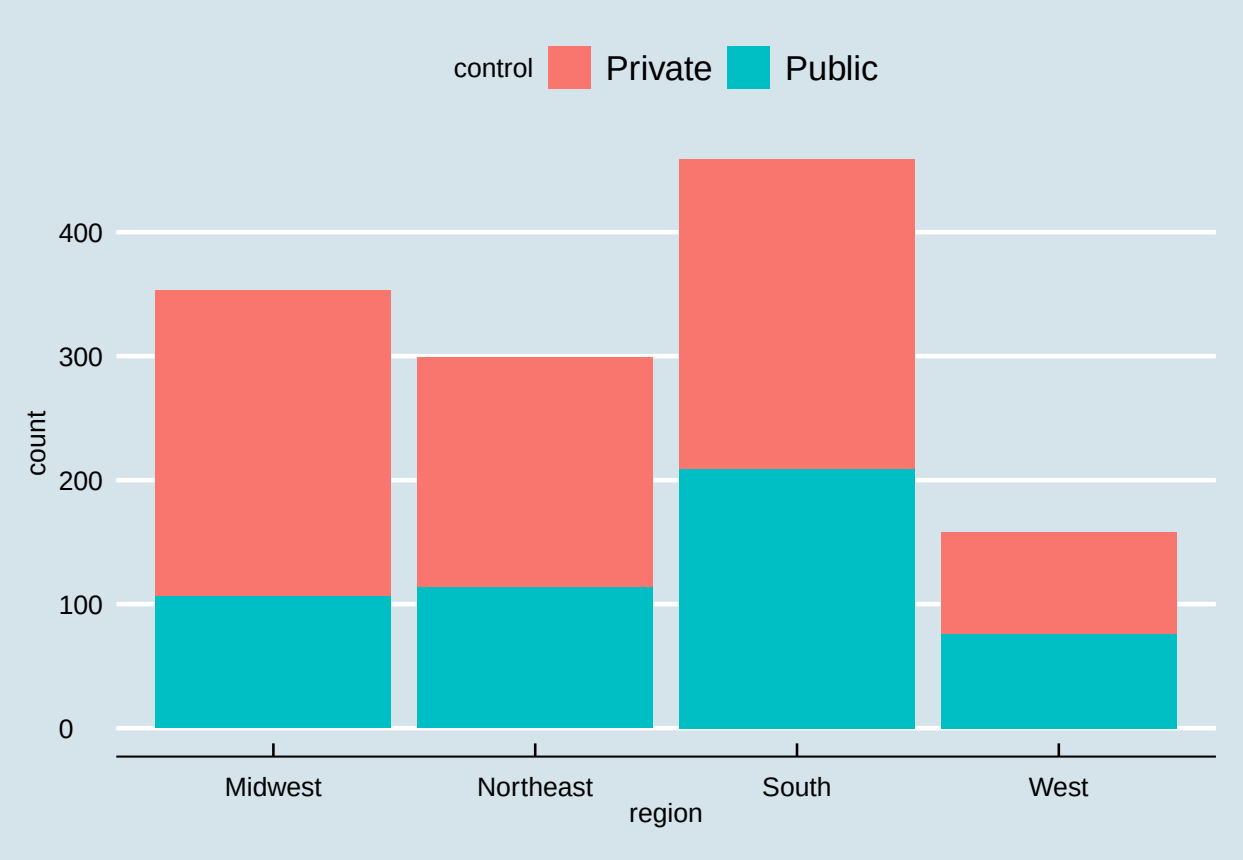
Wsj theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_wsj()
```



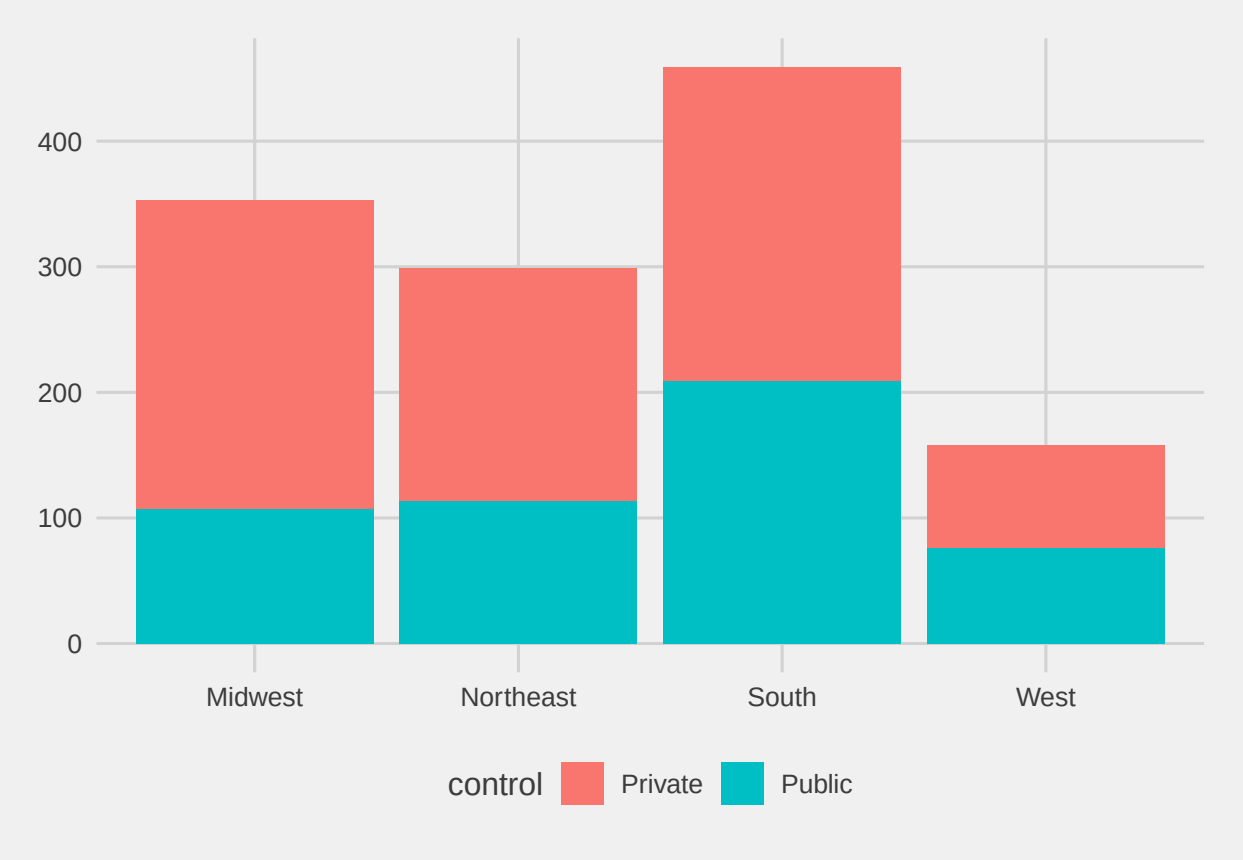
Econimist theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_economist()
```



Fivethirtyeight theme

```
ggplot(data=college) +
  geom_bar(mapping=aes(x=region, fill=control)) +
  theme_fivethirtyeight()
```


400
300
200
100
0
Midwest
Northeast
South
West
control
Private
Public