

HW2

Arjun Ganesh

1. Write a code to determine the number of wooden roller coasters in the State of Virginia.

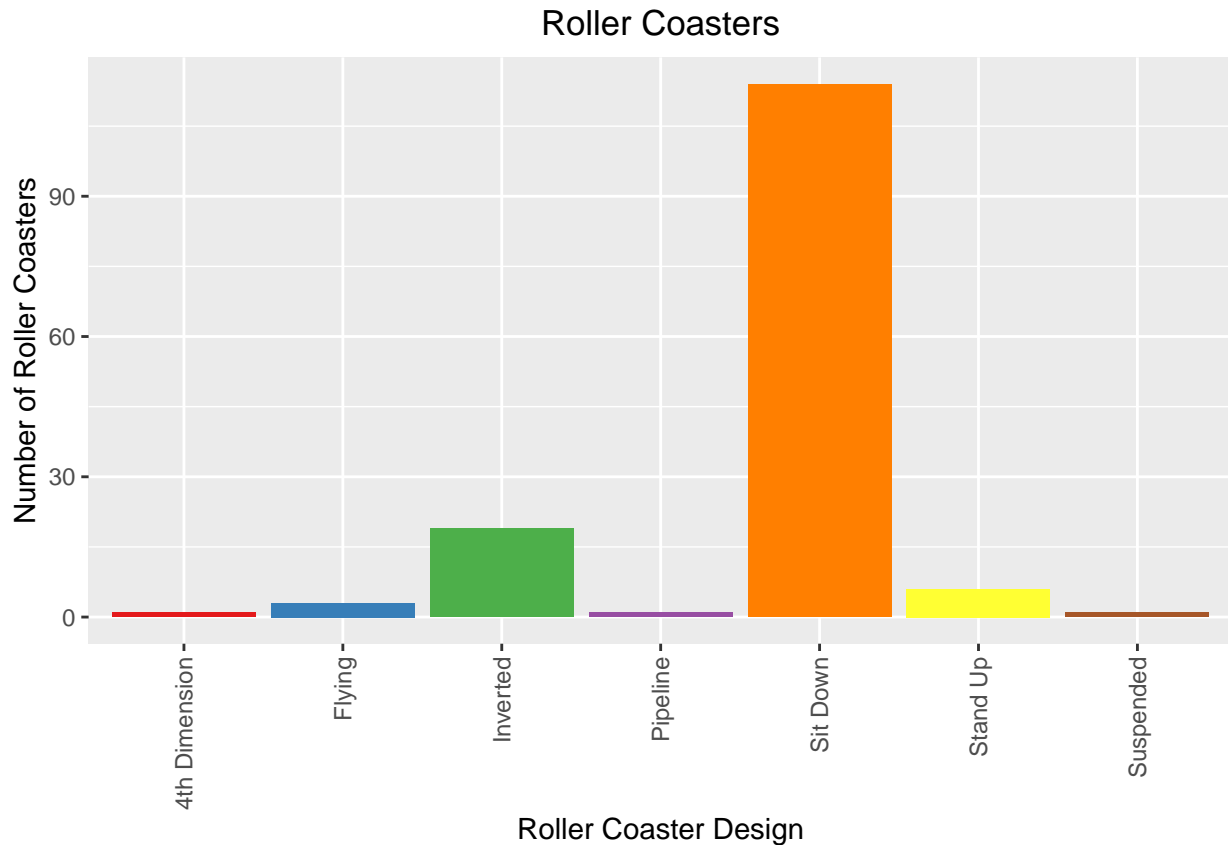
```
rcdata<-read.csv("/cloud/project/RollerCoasterData (1).csv")
virgdata<-subset(rcdata,State=="Virginia")
table(virgdata$Type)
```

```
##
##  Steel  Wooden
##      8      0
```

2. Recreate the graphic exactly as shown below. Also, comment on what this graph reveals.

Based on the graph, it seems that most amount of roller coasters are sit down rollercoasters.

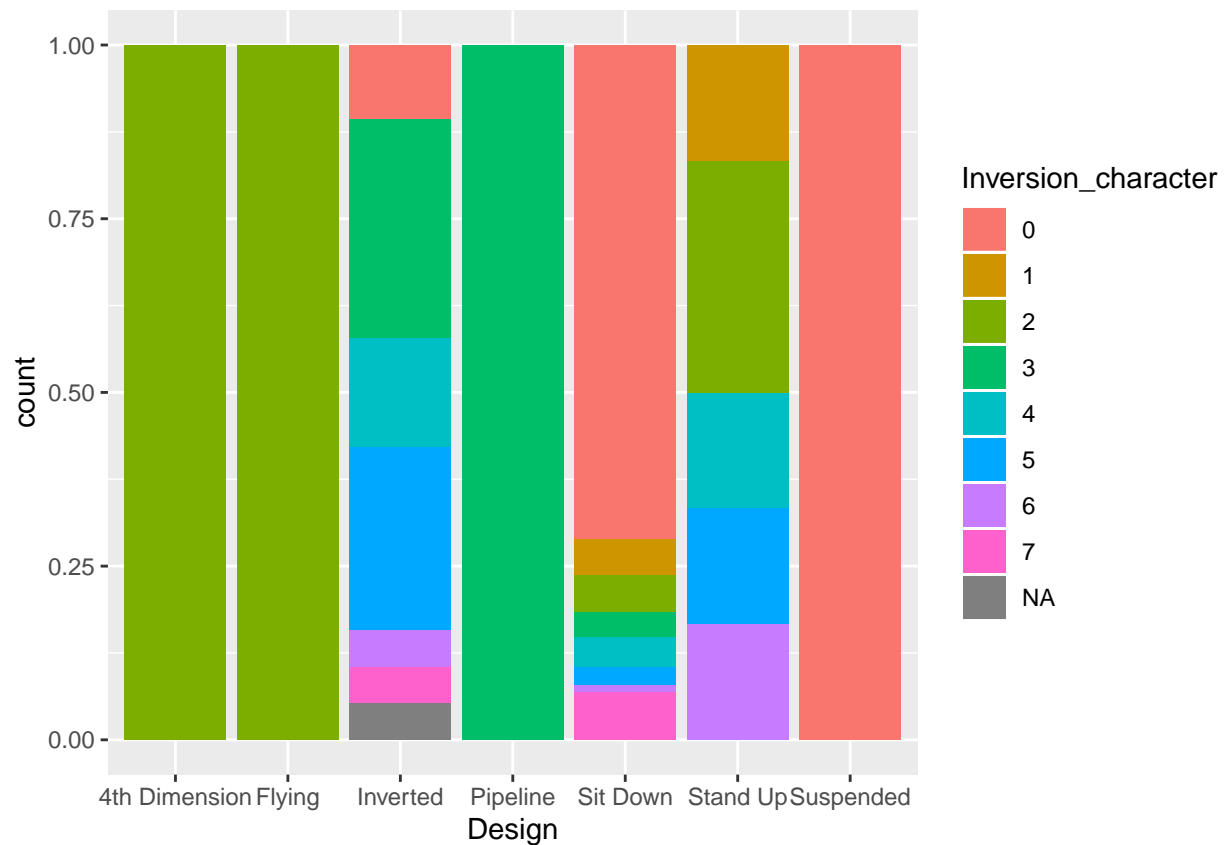
```
rcdata<-read.csv("/cloud/project/RollerCoasterData (1).csv")
library(RColorBrewer)
library(ggplot2)
ggplot(data = rcdata, aes(x =Design)) +
  geom_bar(fill=brewer.pal(7,"Set1"))+
  labs(x = "Roller Coaster Design", y="Number of Roller Coasters",
       title = "Roller Coasters") +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))+
  theme(plot.title = element_text(hjust = 0.5))
```



3. Convert the Inversions variable to a character vector. Make a percent stacked bar graph to visualize the converted Inversions variable and the Design variable. Also, use your graph to comment on whether the design of a roller coaster is independent of the number of inversions.

I would say the design of roller coaster is dependent of the number of inversions based on the graph. From further research, certain designs allow for inversions. For example, suspended types cannot have any inversions. That is why I came to this conclusion.

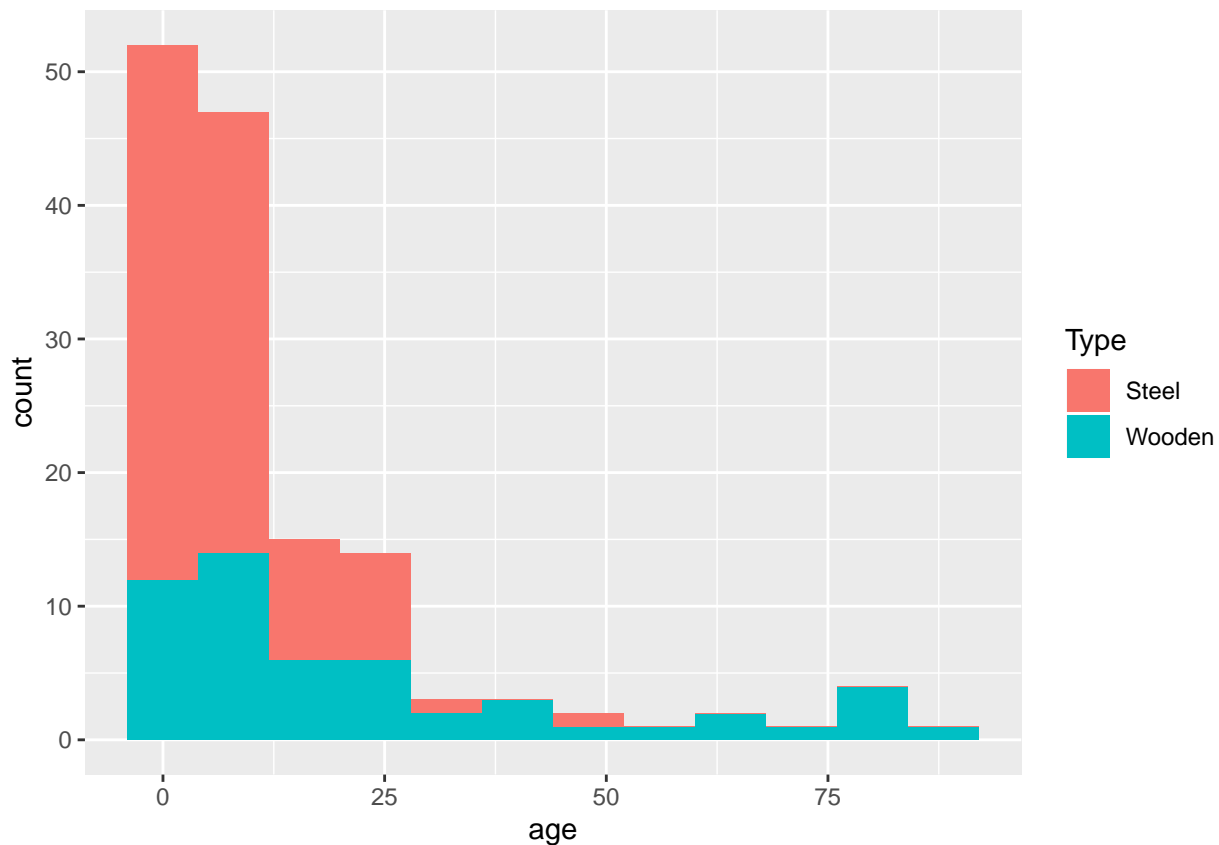
```
Inversions<-rcdata$Inversions
Inversion_character<-as.character(Inversions)
ggplot(data = rcdata, mapping = aes(x =rcdata$Design,fill= Inversion_character)) +
  geom_bar(position="fill")+xlab("Design")
```



4. Recreate the graphic exactly as shown below. Also, comment on what this graph reveals.

This graph reveals that there are many new and young roller coasters. And the reason for such low count of older ones is clearly due to structural issues and different codes related to the safety of riding a rollercoaster.

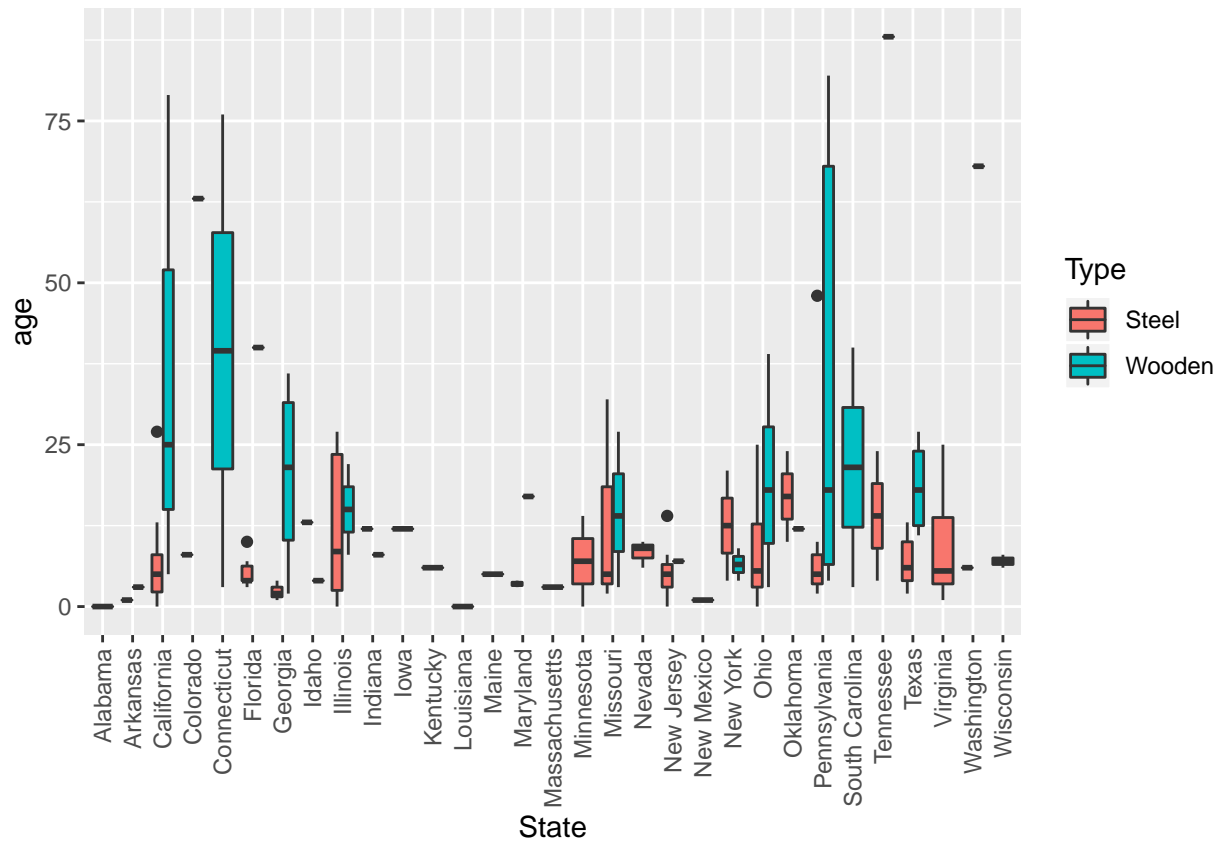
```
library(ggplot2)
ggplot(rcdata, aes(x=age, fill=Type)) +
  geom_histogram(bins=12)
```



5. Recreate the graphic exactly as shown below. Use the graph to determine the state with the oldest roller coaster. Also, determine the type (wooden or steel) of the oldest roller coaster.

According to the boxplot, the oldest roller coaster seems to be wooden and located in Tennessee. The data point is extremely far away from the boxplot.

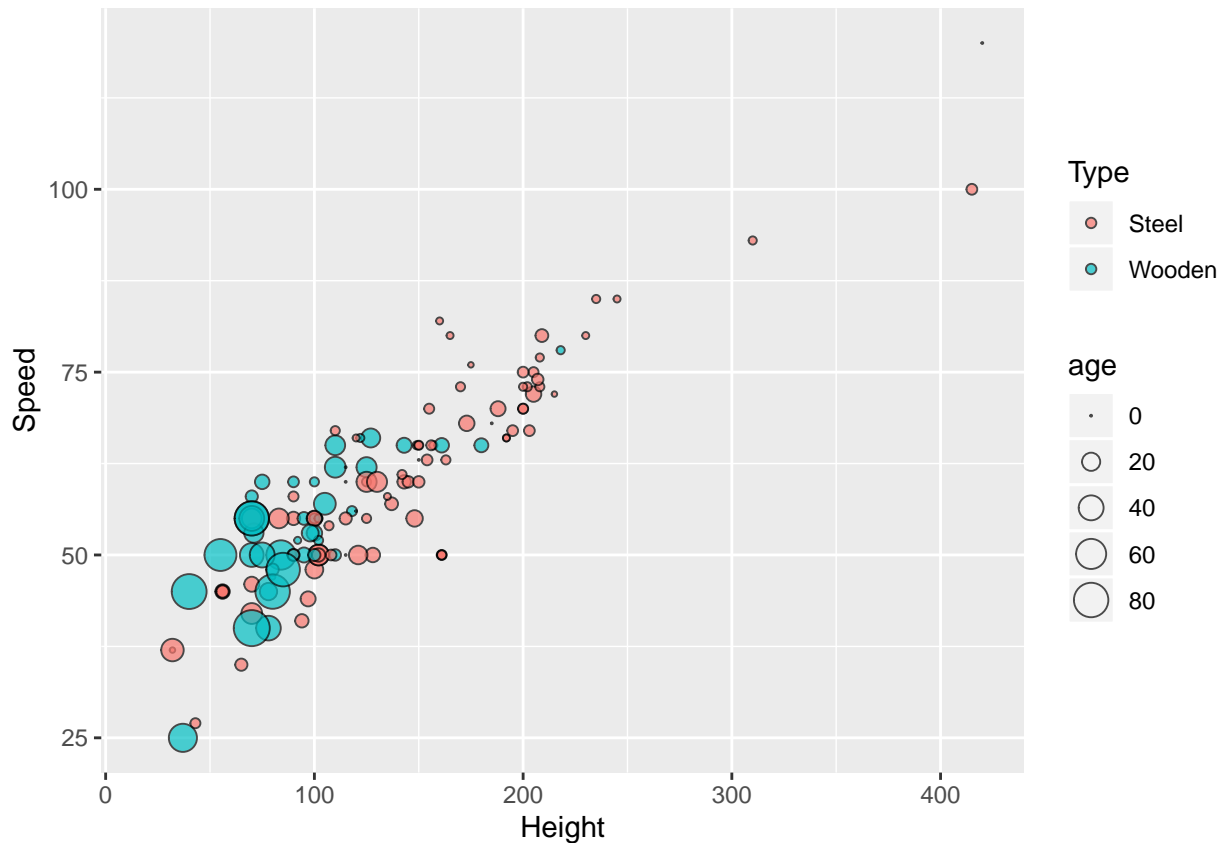
```
ggplot(data = rcddata, mapping = aes(x=State,y=age,fill=Type)) +  
  geom_boxplot()+  
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))
```



6.Recreate the graphic similarly as shown below. Also, comment on what this graph reveals.

A generalized comment that can be made from the graphic is that the taller the rollercoaster it seems the faster it rides. Also, wooden coasters are slower than steel coasters.

```
ggplot(rcdata, aes(x=Height,y=Speed, size= age, fill = Type))+
  geom_point(alpha=0.7,shape=21,color="black",na.rm = TRUE)+
  scale_size(range=c(0,6))
```



7. Open-ended: use some the variables in this dataset to create a visual display of interest to you. Why did you choose the variables and this visualization? Also, comment on what this graph reveals.

I chose to do a scatter plot but for each type of design. The three other variables compared are Speed and Drop as the main exploratory and reasoning variables. I simply added Height to add some extra fun. It seems that Sit down rollercoasters have medium height but very spread out on the how fast they ride with a higher drop rate.

```
ggplot(rcdata,aes(x=Drop,y=Speed,colour=Height))+geom_point()+facet_wrap(~Design)
```

```
## Warning: Removed 13 rows containing missing values (geom_point).
```

