Introduction to R Markdown

2018-07-11

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

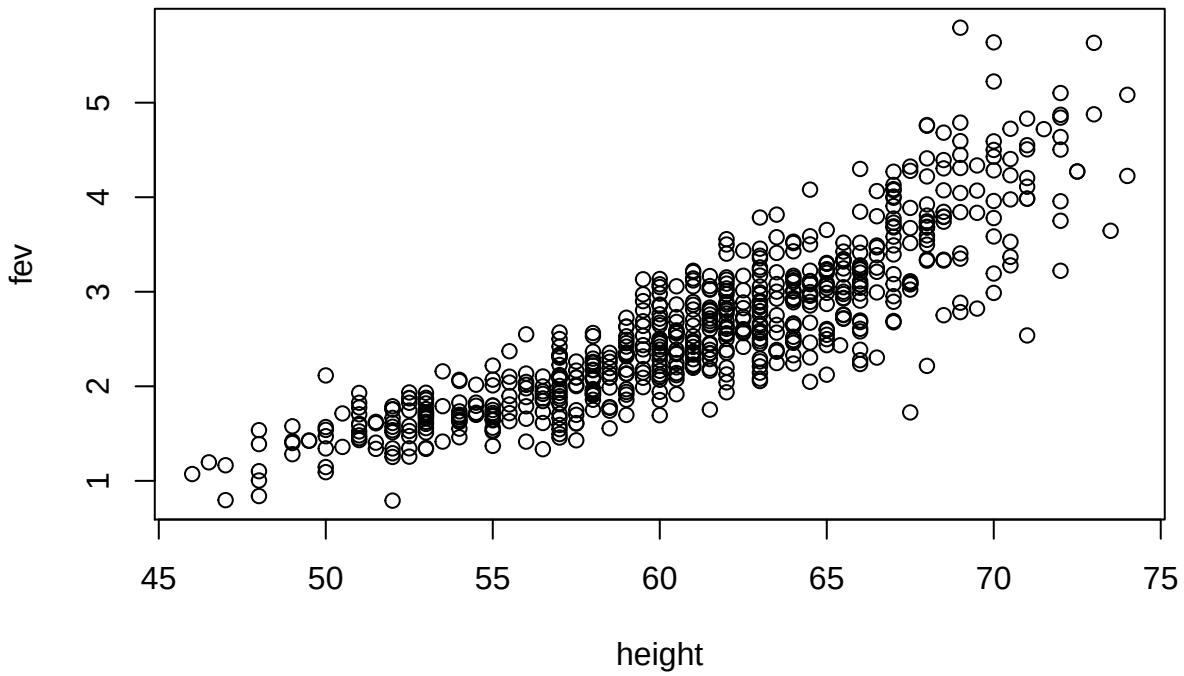
A quick references to the most commonly used R Markdown syntax can be found here: http://rmarkdown.rstudio.com/authoring_basics.html

An extensive R Markdown cheatsheet can be found here: https://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf

When you click the **Knit** button in Rstudio, a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

In an R Markdown file you type the R code that you want in "chunks" as follows:

```
fevdata <- read.csv("http://faculty.washington.edu/tathornt/Biost509/DataSets/fev2.csv",header=TRUE)
plot(fev ~ height, data=fevdata)
```



Note the first and last lines of the chunk. These are required to show when a chunk begins and ends. The R code follows the same syntax in R markdown as in an R script file.

To run a chunk of R code, place your cursor anywhere in the chunk, click the "Chunks" button and select "Run Current Chunk".

To execute the entire file, click the arrow to the right of the "Knit" button and select your desired output format (pdf, html or word). A document will be generated that includes the text content outside of the chunks as well as the output of the R code chunks.

The files you create will open automatically and will also be saved in the working directory.

Look at the first few lines of the file and obtain summary statistics for each variable

```
head(fevdata)
```

```
##   seqnbr subjid age   fev height sex smoke
## 1      1    301   9 1.708   57.0   2     2
## 2      2    451   8 1.724   67.5   2     2
## 3      3    501   7 1.720   54.5   2     2
## 4      4    642   9 1.558   53.0   1     2
## 5      5    901   9 1.895   57.0   1     2
## 6      6   1701   8 2.336   61.0   2     2
```

```
summary(fevdata)
```

```
##      seqnbr          subjid           age              fev
##  Min.   :  1.0   Min.   :  201   Min.   : 3.000   Min.   :0.791
##  1st Qu.:164.2   1st Qu.:15811   1st Qu.: 8.000   1st Qu.:1.981
##  Median :327.5   Median :36071   Median :10.000   Median :2.547
##  Mean   :327.5   Mean   :37170   Mean   : 9.931   Mean   :2.637
##  3rd Qu.:490.8   3rd Qu.:53638   3rd Qu.:12.000   3rd Qu.:3.119
##  Max.   :654.0   Max.   :90001   Max.   :19.000   Max.   :5.793
##      height           sex            smoke
##  Min.   :46.00   Min.   :1.000   Min.   :1.000
##  1st Qu.:57.00   1st Qu.:1.000   1st Qu.:2.000
##  Median :61.50   Median :1.000   Median :2.000
##  Mean   :61.14   Mean   :1.486   Mean   :1.901
##  3rd Qu.:65.50   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :74.00   Max.   :2.000   Max.   :2.000
```

Create a new sex variable for males and females

```
fevdata$sex2<-ifelse(fevdata$sex==1,"male","female")
fevdata$sex2<-as.factor(fevdata$sex2)
```

Obtain the mean fev for males and females

```
avgmalefev<-mean(fevdata$fev[fevdata$sex2=="male"],na.rm=TRUE)
avgmalefev
```

```
## [1] 2.812446
```

```
avgfemalefev<-mean(fevdata$fev[fevdata$sex2=="female"],na.rm=TRUE)
avgfemalefev
```

```
## [1] 2.45117
```

Boxplots of FEV for males and females.

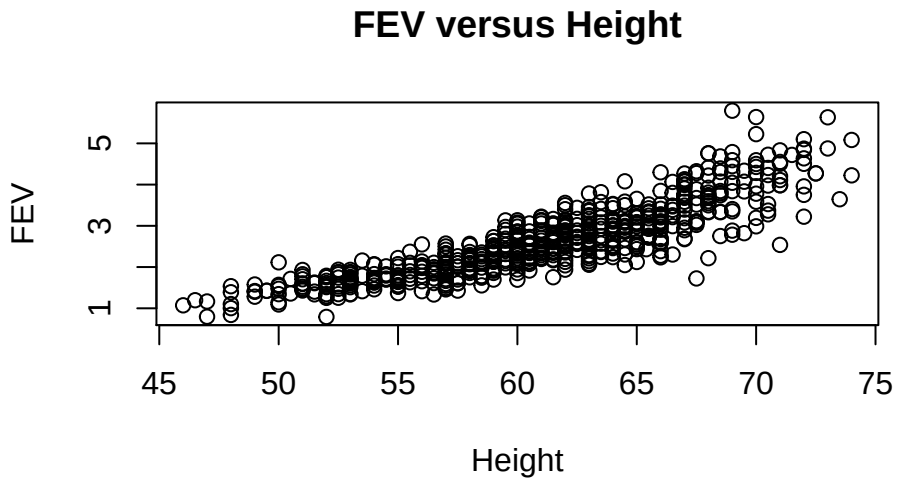
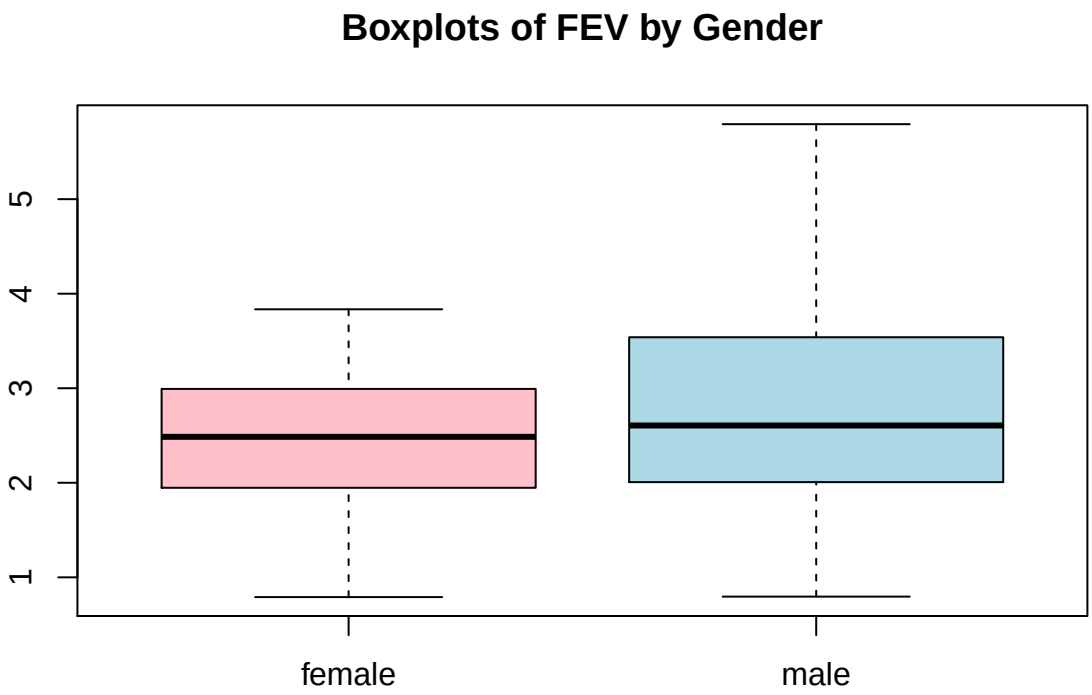


Figure 1: Scatterplot and Regression Line of of FEV on Height

```
### Box plot of FEV for males and females ###

boxplot(fev ~ sex2,data=fevdata,col=c("pink","lightblue"),main="Boxplots of FEV by Gender")
```



For a page or line break in the document, use three or more astericks (*) or dashes (-).

Figure dimensions are controlled by the fig.height and fig.width parameters (units are inches). You can also add a caption with the fig.cap parameter.

```
plot(fev ~ height,ylab="FEV", xlab="Height",main="FEV versus Height",data=fevdata)
```

Note that for the following chunk, the R code is suppressed by the "echo=F" paramater. This is useful when you are using R markdown to write a report and only want to see the results. Here is a plot, without the R

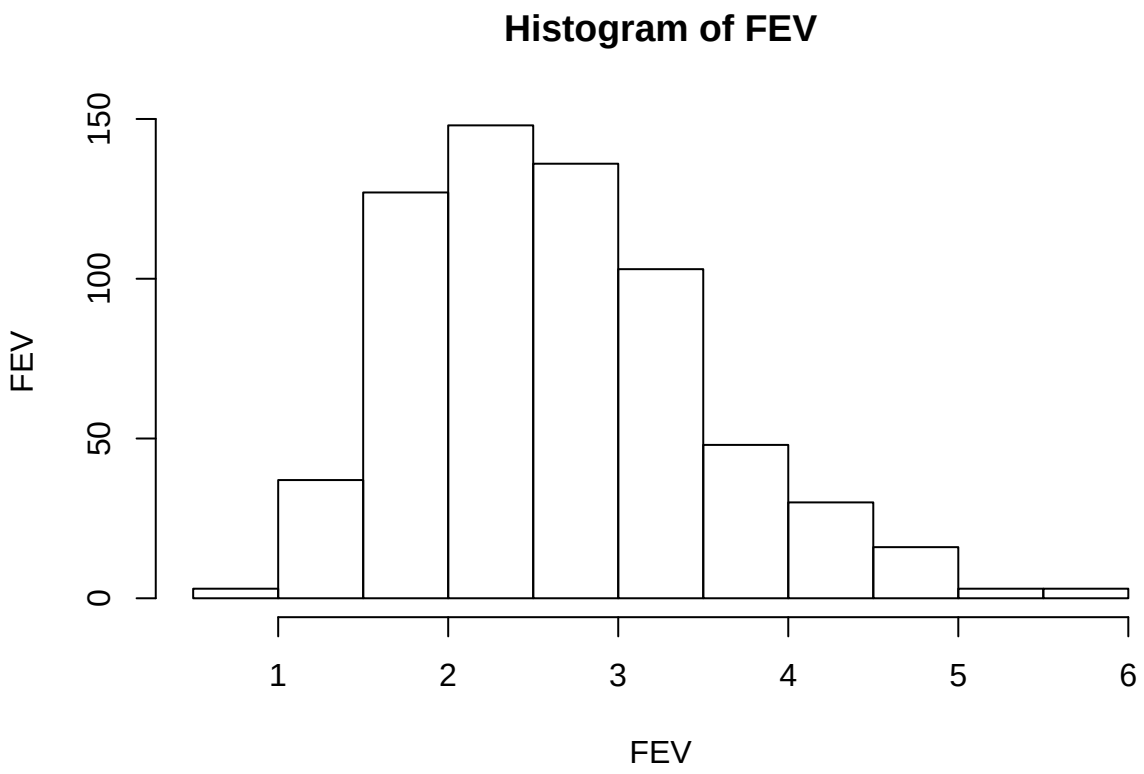


Figure 2: Histogram of FEV

code appearing in the document:

Extension R packages can easily be used with Rmarkdown.

For example, can use the **ggplot2** package for data visualization.

```
library(ggplot2)
```

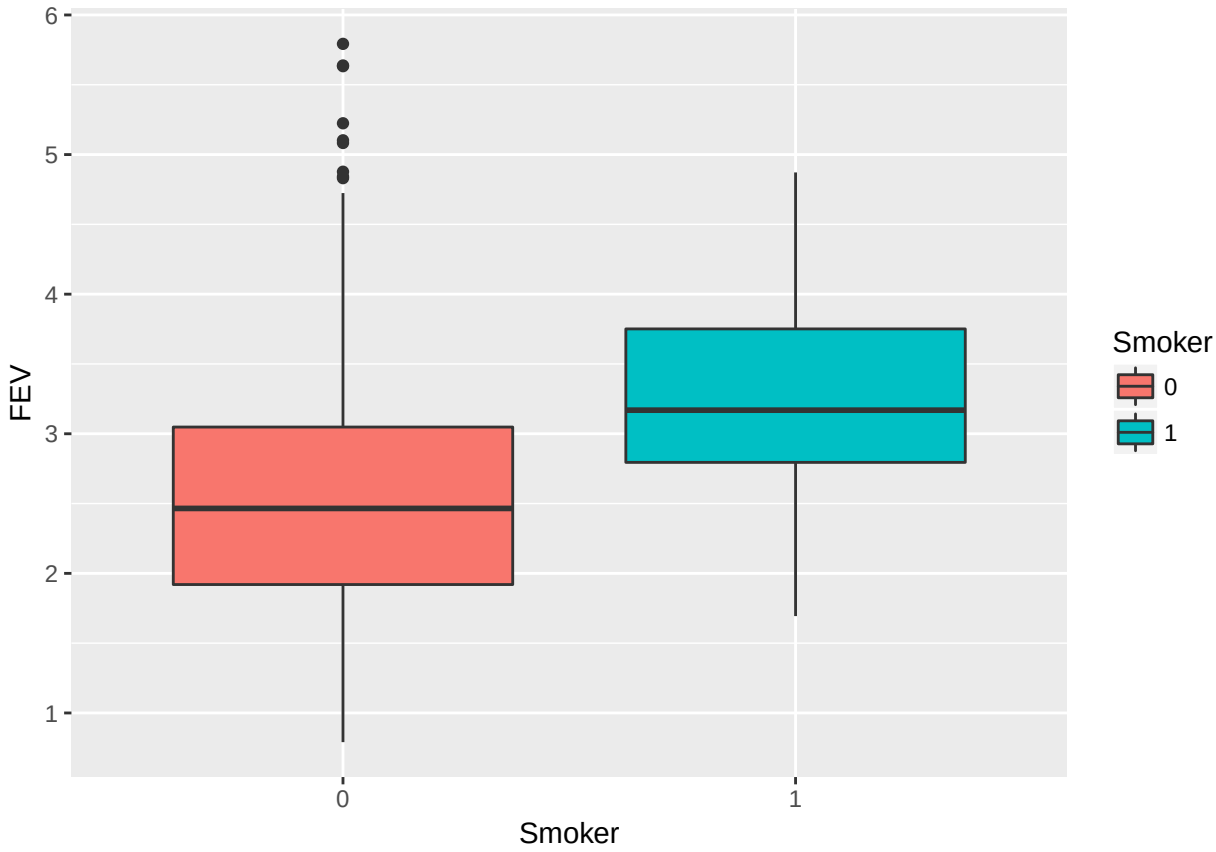
Suppose we are interested in investigating the relationship between smoking and FEV. Let's first create a new smoking indicator variable for smoker, where 1 corresponds to a smoker and 0 corresponds to a non-smoker

```
fevdata$smoker<-(2-fevdata$smoke)
head(fevdata)
```

```
##   seqnbr subjid age   fev height sex smoke   sex2 smoker
## 1      1    301   9 1.708   57.0   2     2 female      0
## 2      2    451   8 1.724   67.5   2     2 female      0
## 3      3    501   7 1.720   54.5   2     2 female      0
## 4      4    642   9 1.558   53.0   1     2   male      0
## 5      5    901   9 1.895   57.0   1     2   male      0
## 6      6   1701   8 2.336   61.0   2     2 female      0
```

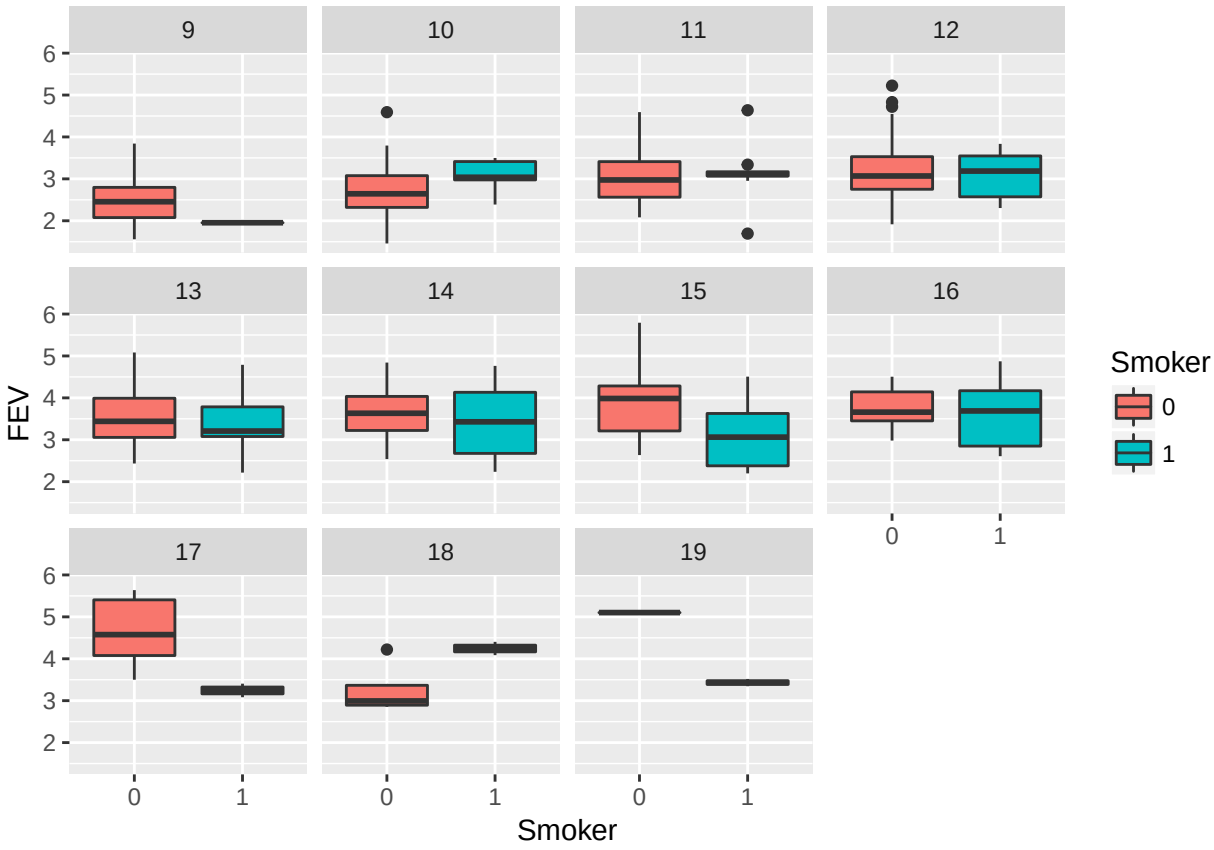
Boxplot of FEV by smoking group

```
ggplot(fevdata,aes(x=as.factor(smoker),y=fev,fill=as.factor(smoker)))+ geom_boxplot() +xlab("Smoker") +
```



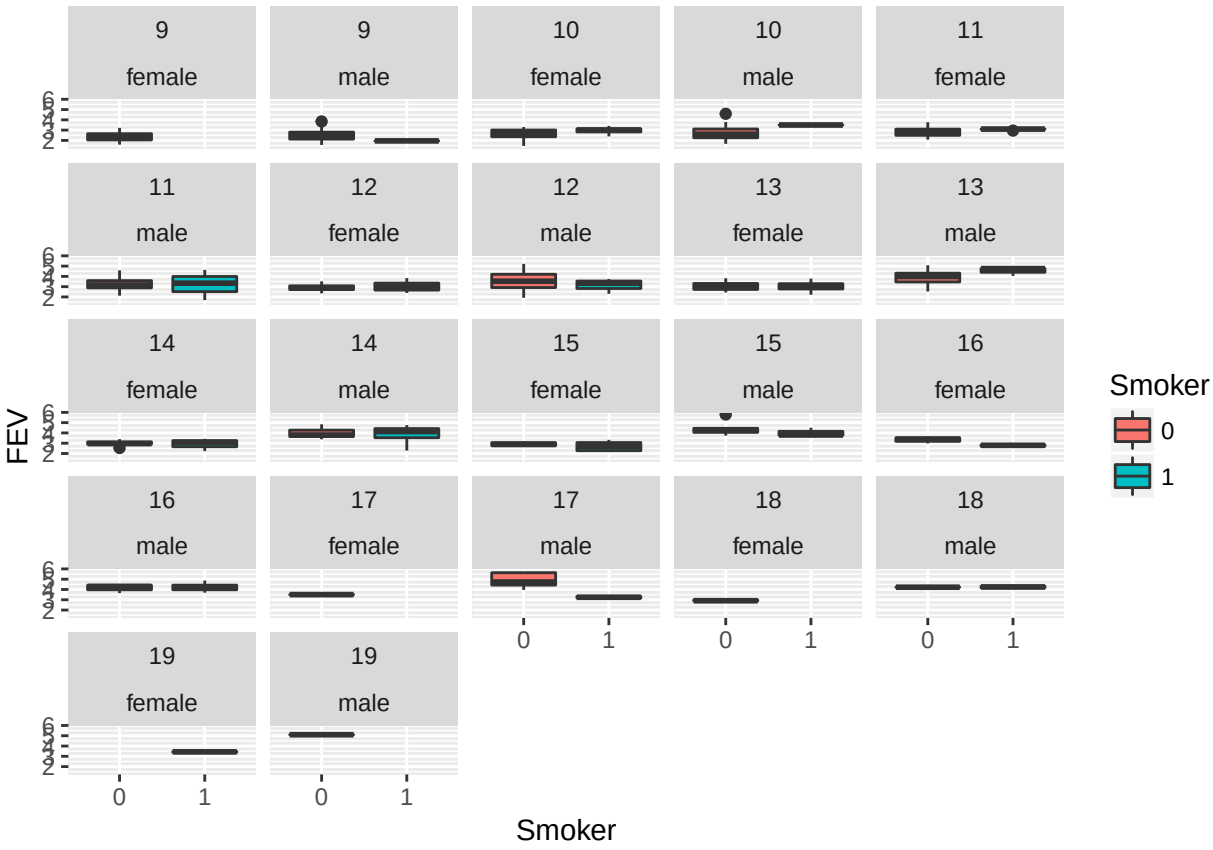
Boxplot of FEV by smoking group across each age group that has both non-smokers and smokers

```
fevdata2<-subset(fevdata,age>=9)
ggplot(fevdata2,aes(x=as.factor(smoker),y=fev,fill=as.factor(smoker)))+ geom_boxplot() +xlab("Smoker") +
```



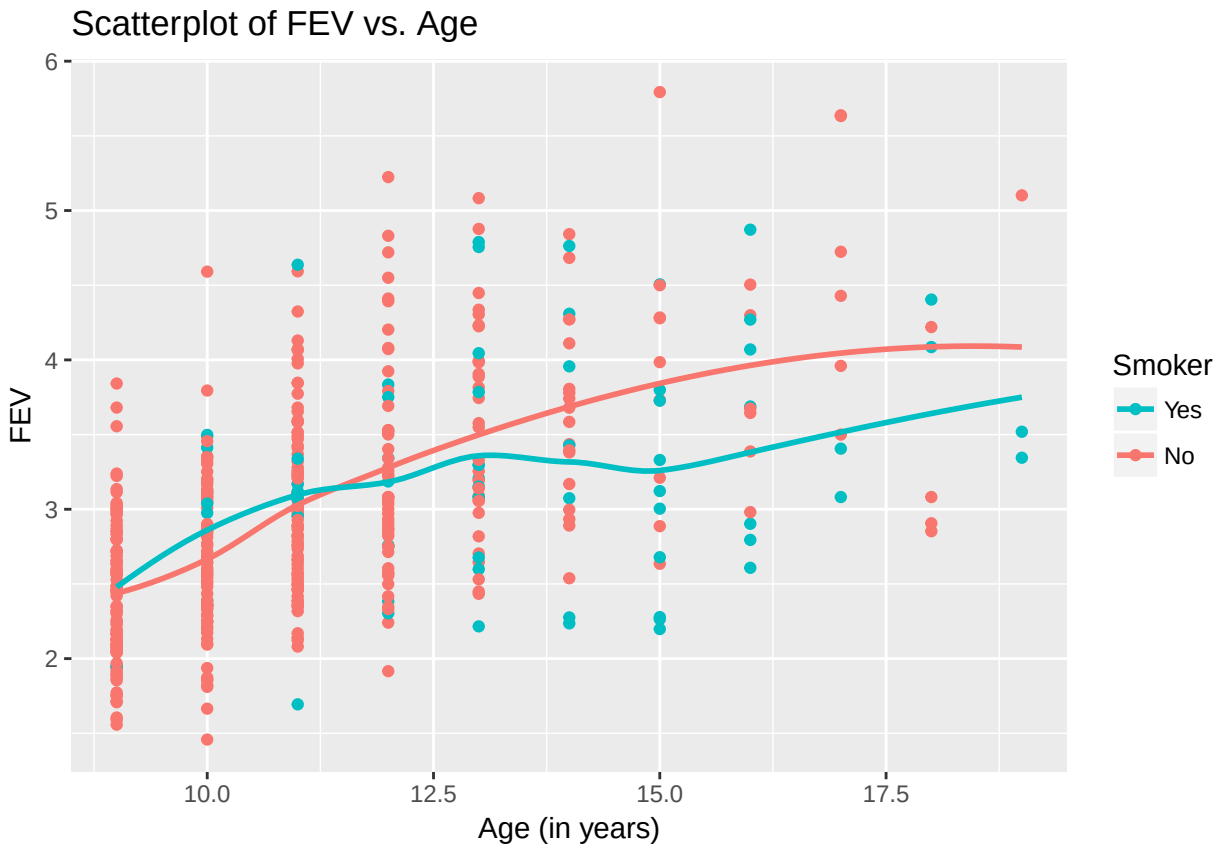
Boxplot of FEV by smoking group across each age and gender group that has both non-smokers and smokers

```
ggplot(fevdata2,aes(x=as.factor(smoker),y=fev,fill=as.factor(smoker)))+ geom_boxplot() +xlab("Smoker")
```



Scatterplot of FEV by age with LOESS smoothing curve for each smoking group

```
p<-ggplot(fevdata2,aes(x=age,y=fev,colour=as.factor(smoker)))
p+geom_point(size=1.5)+geom_smooth(method="loess",se=FALSE)+xlab("Age (in years)")+ylab("FEV")+scale_co
```


Scatterplot of FEV vs. Age
FEV
6
5
4
3
2
10.0
12.5
15.0
17.5
Age (in years)
Smoker
Yes
No