Assignment1

2023-07-31

These exercises require you to generate plots of various kinds.

Part I – geoms and aesthetics

These first few exercises will run through some of the simple principles of creating a ggplot2 object, assigning aesthetics mappings and geoms.

1. Read in the cleaned patients dataset, `patient-data-cleaned.txt`, into a new object called `patients`.

```
library(ggplot2)

patients = read.table("patient-data-cleaned.txt", header=T, sep="\t", fill=T)

head(patients)
```

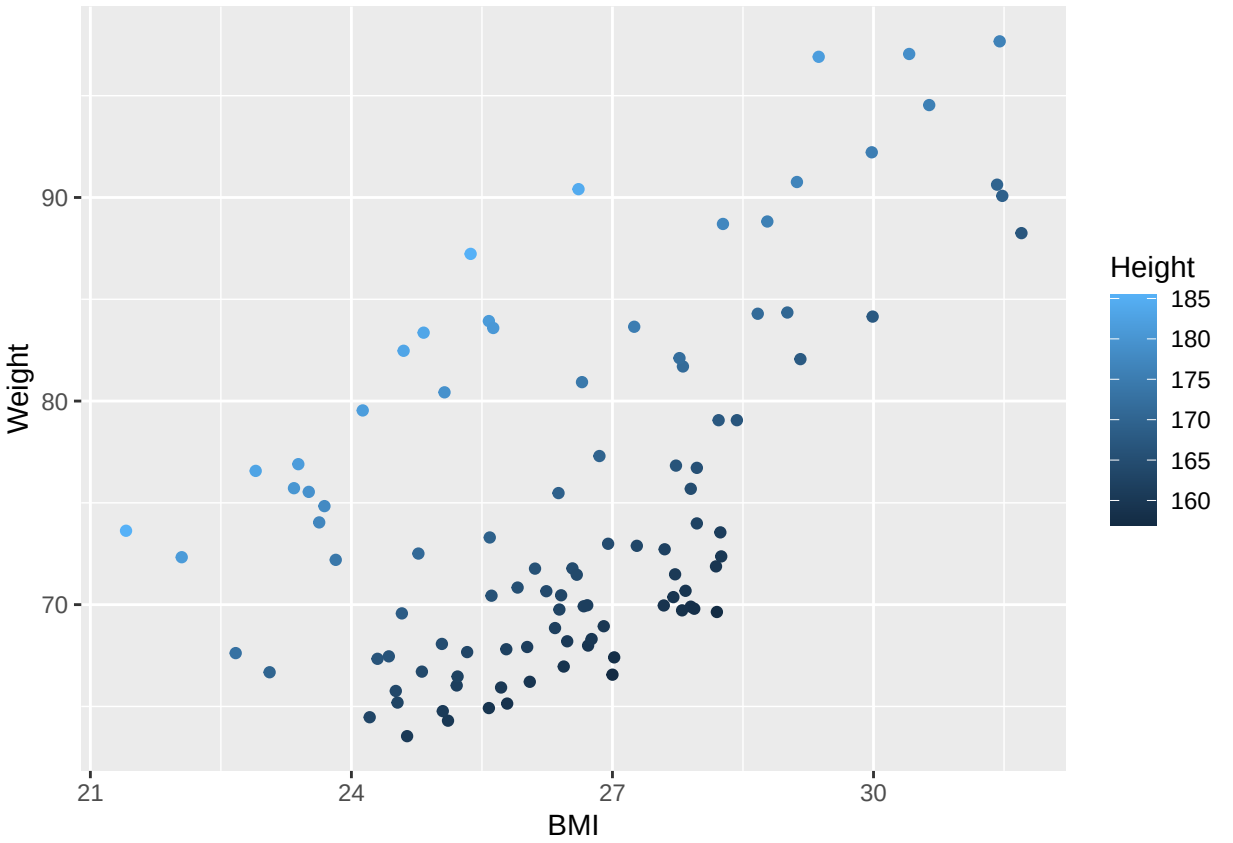
```
##          ID      Name    Sex     Smokes Height Weight      Birth      State
## 1 AC/AH/001   Michael   Male Non-Smoker 182.87  76.57 1972-02-06    Georgia
## 2 AC/AH/017     Derek   Male Non-Smoker 179.12  80.43 1972-06-15   Colorado
## 3 AC/AH/020      Todd   Male Non-Smoker 169.15  75.48 1972-07-09 New Jersey
## 4 AC/AH/022    Ronald   Male Non-Smoker 175.66  94.54 1972-08-17   Colorado
## 5 AC/AH/029 Christine Female Non-Smoker 164.47  71.78 1973-06-12    Georgia
## 6 AC/AH/033      Dana Female     Smoker 158.27  69.90 1973-07-01    Indiana
##   Grade  Died Score Date.Entered.Study Age   BMI Overweight
## 1     2 FALSE  0.01         2015-12-01  44 22.90      FALSE
## 2     2 FALSE -1.31         2015-12-01  43 25.07       TRUE
## 3     2 FALSE -0.17         2015-12-01  43 26.38       TRUE
## 4     1 FALSE -1.10         2015-12-01  43 30.64       TRUE
## 5     2  TRUE  1.42         2015-12-01  42 26.54       TRUE
## 6     2 FALSE  0.29         2015-12-01  42 27.90       TRUE
```

Scatterplots

2. Generate a scatterplot of BMI versus Weight using the patient dataset and add a colour scale based on the Height variable.

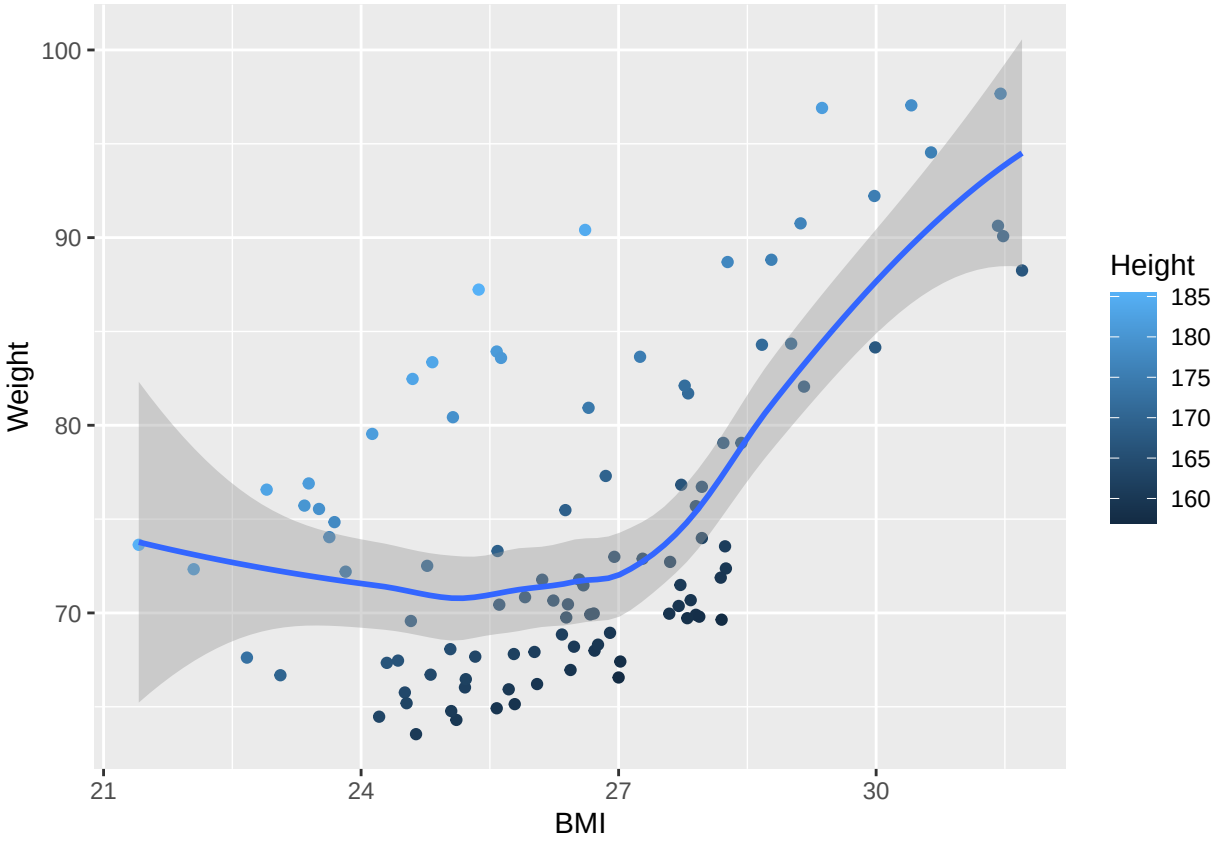
```
s = ggplot(data = patients, mapping = aes(x = BMI, y = Weight)) +
  geom_point(mapping = aes(color = Height))

s
```



3. Using an additional geom, add an extra layer of a fit line to the previous plot.(use geom_smooth())

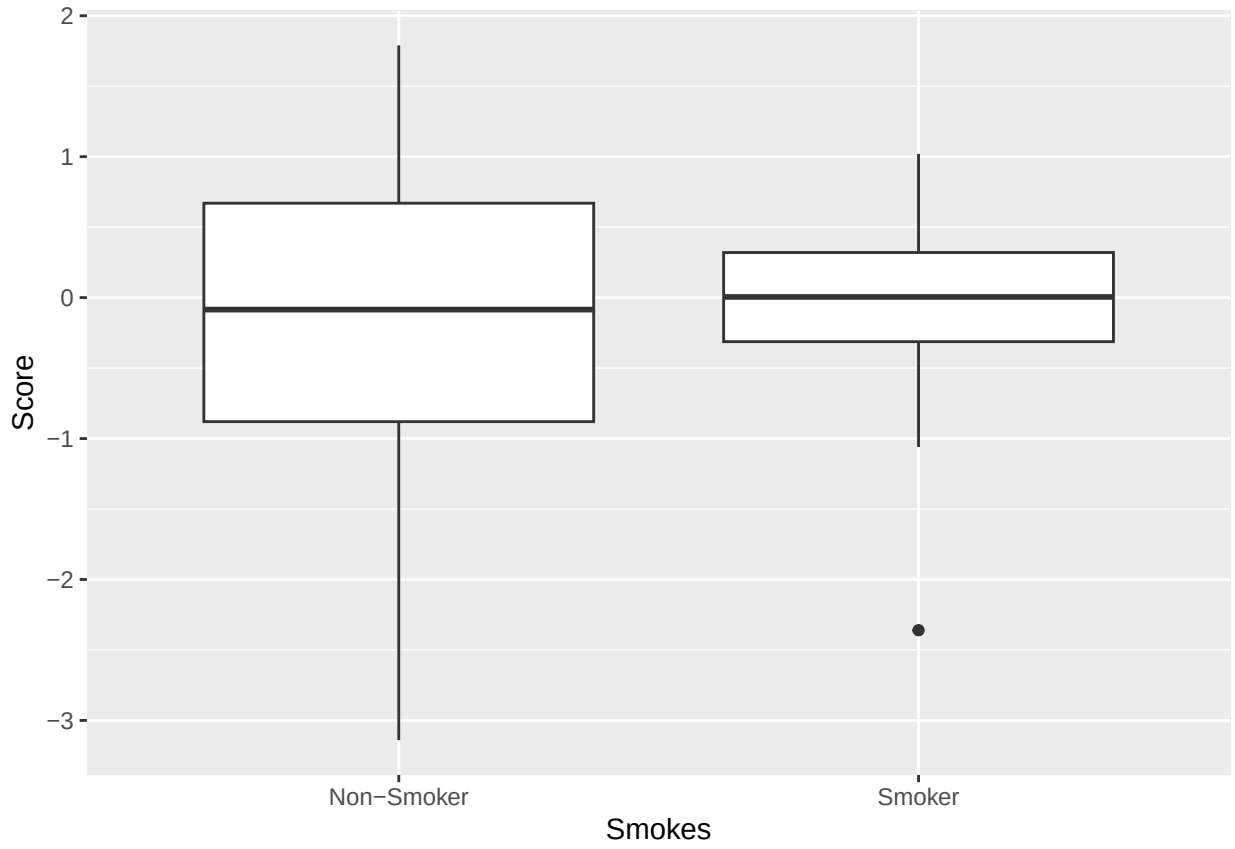
```
s + geom_smooth()
```



Boxplots and Violin plots

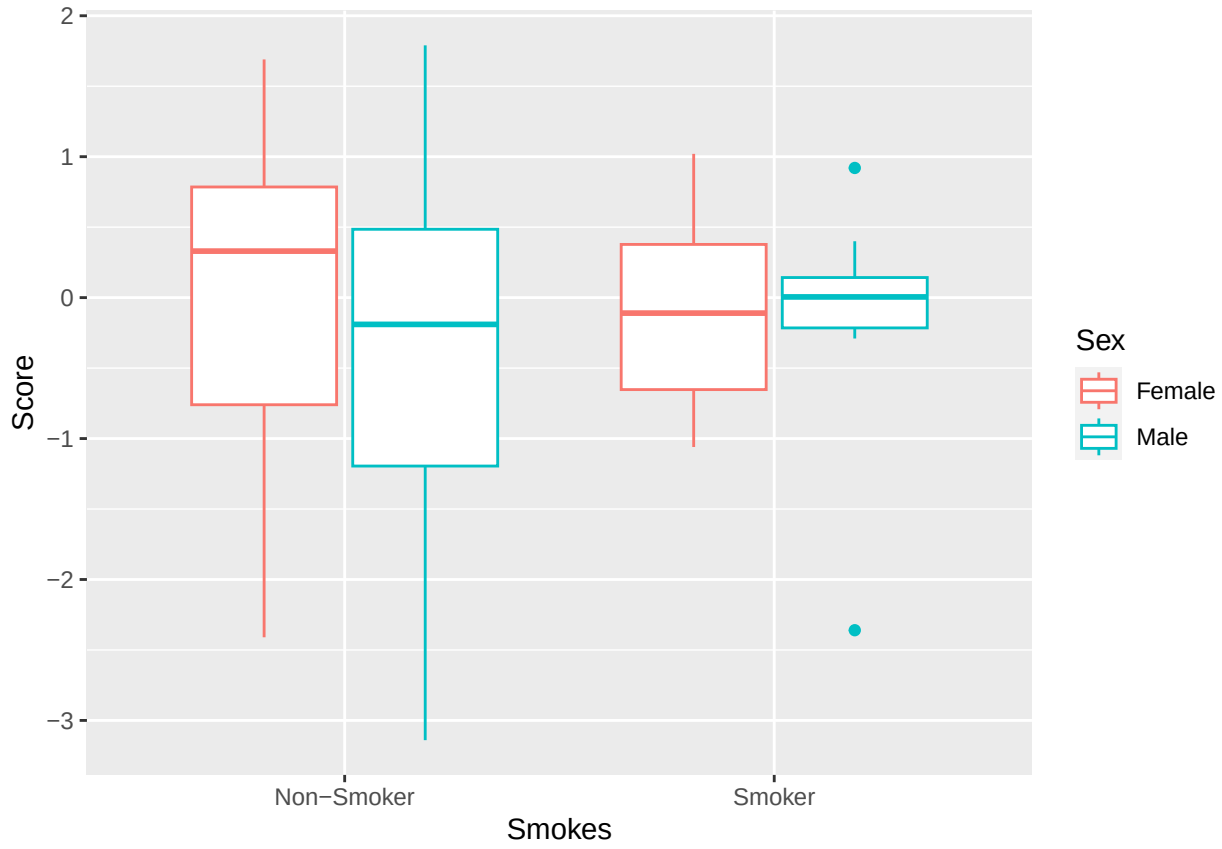
4. Generate a boxplot of Score values comparing smokers and non-smokers.

```
ggplot(data = patients, mapping = aes(x = Smokes, y = Score)) +
  geom_boxplot()
```



5. Split the previous boxplot into male and female groups with different colours.

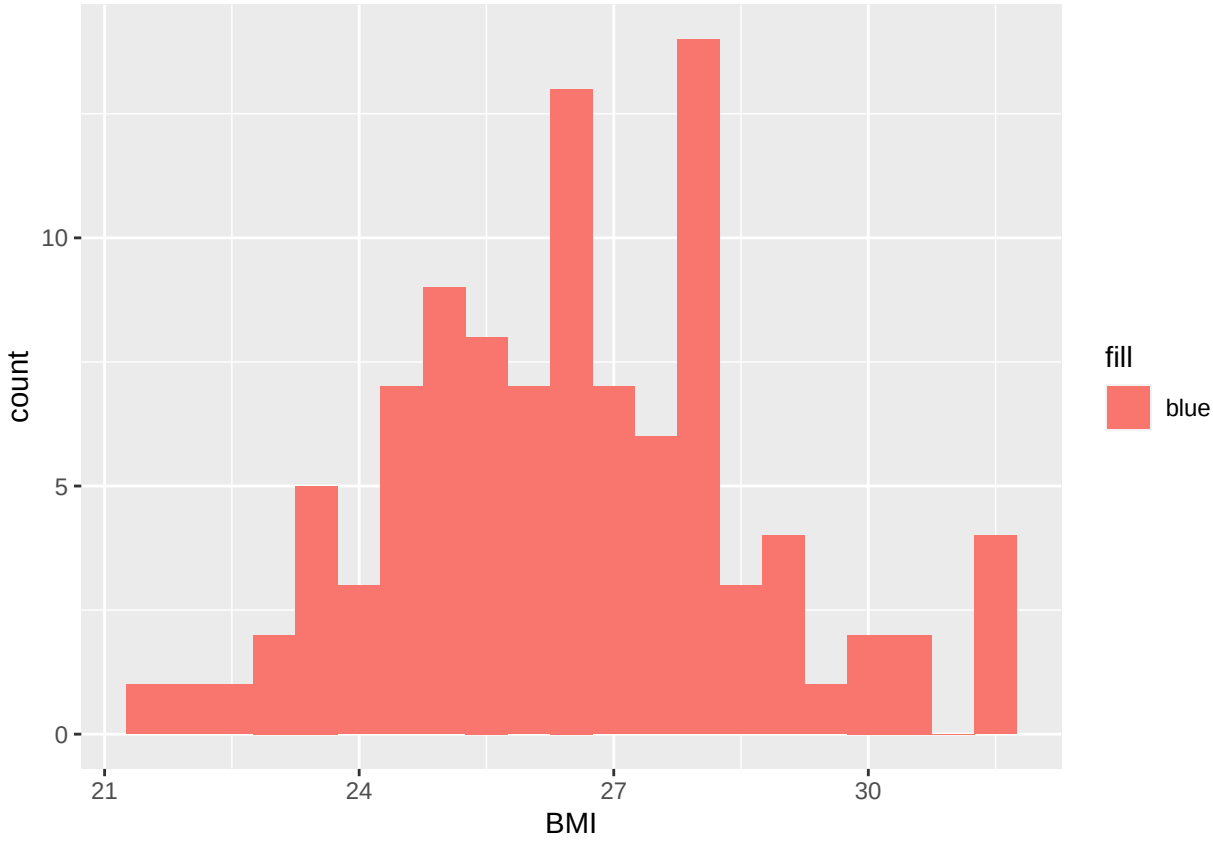
```
ggplot(data = patients, mapping = aes(x = Smokes, y = Score)) +
  geom_boxplot(mapping = aes(color = Sex))
```



Histogram and Density plots

6. Generate a histogram of BMIs with each bar coloured blue, choosing a suitable bin width.

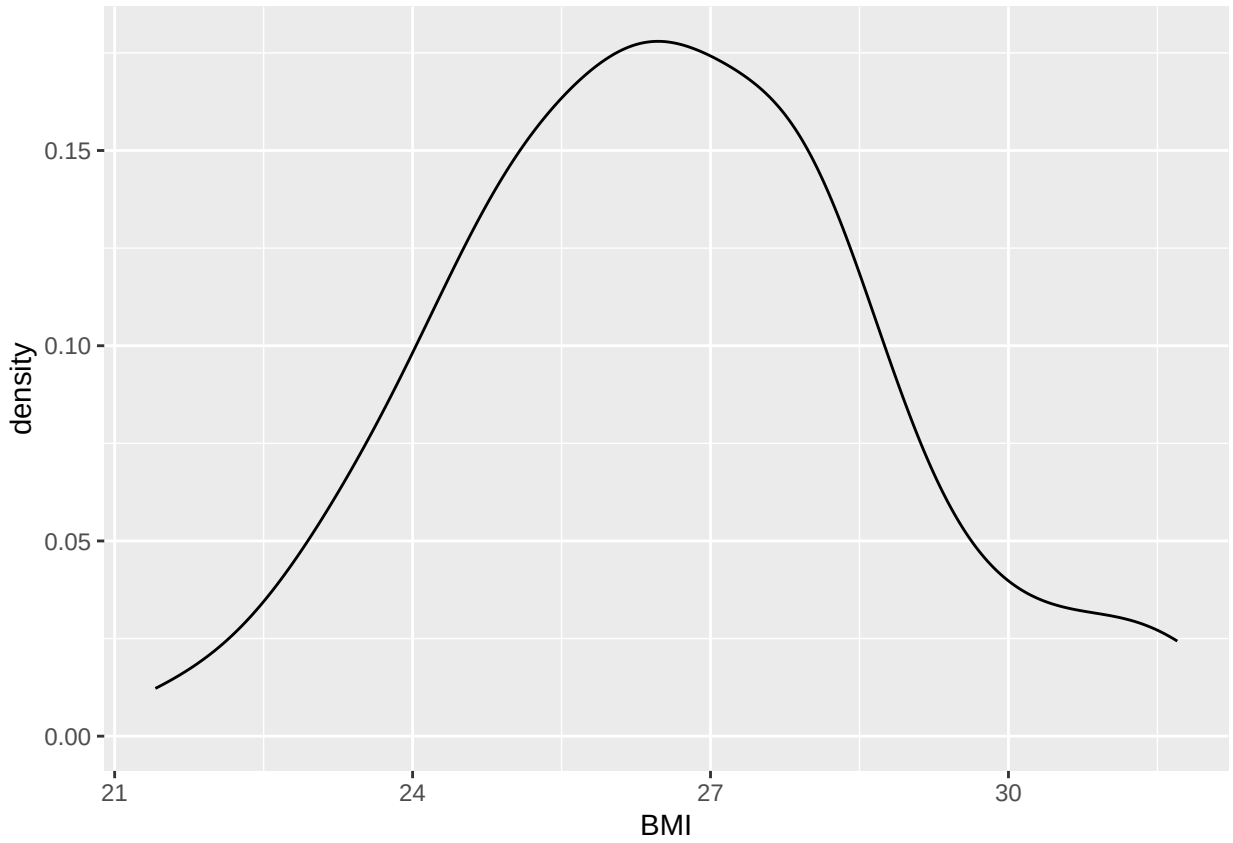
```
ggplot(data = patients, mapping = aes(x = BMI)) +
  geom_histogram(binwidth = 0.5, mapping = aes(fill = "blue"))
```




```
# I do not why it outputs as red
```

7. Instead of a histogram, generate a density plot of BMI

```
ggplot(data = patients, mapping = aes(x = BMI)) +
  geom_density()
```


0.15
0.10
0.05
0.00
density
21
24
27
30
BMI