Graphics Self-Test
SER 2017 R Workshop

S Mooney

June 19, 2017

1. Load the digitalis dataset (the url is `http://www.columbia.edu/~sjm2186/EPIC_R/dig.csv`) into a data frame named dig. Make sure any character fields load as characters, not as factors

```
> dig <- read.csv("http://www.columbia.edu/~sjm2186/EPIC_R/dig.csv",
+                   stringsAsFactors=F)
>
```

2. How many columns of data are in dig? How many rows?

```
> str(dig)

'data.frame':        6800 obs. of  72 variables:
 $ ID      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ TRTMT   : int  0 0 0 1 0 0 1 1 0 1 ...
 $ AGE     : int  66 77 72 57 74 69 64 60 74 64 ...
 $ RACE    : int  1 1 1 1 1 2 1 2 2 1 ...
 $ SEX     : int  1 1 2 1 1 2 2 1 1 2 ...
 $ EJF_PER : int  40 12 36 31 15 45 30 39 33 24 ...
 $ EJFMETH : int  2 1 1 1 1 1 1 1 3 1 ...
 $ CHESTX  : num  0.5 0.56 0.68 0.48 0.53 0.7 0.52 0.4 0.49 0.52 ...
 $ BMI     : num  20.1 20.7 25.5 25.8 25.7 ...
 $ KLEVEL  : num  NA 3.1 5.1 NA 4 4.3 4.3 5.1 4.7 4 ...
 $ CREAT   : num  1.26 1.5 1.24 2.25 1.47 ...
 $ DIGDOSER: num  0.25 0.25 0.25 0.25 0.375 0.25 0.25 0.25 0.25 0.25 ...
 $ CHFDUR  : int  96 2 12 24 288 84 31 6 12 33 ...
 $ RALES   : int  0 0 2 2 2 2 0 0 2 2 ...
 $ ELEVJVP : int  0 0 0 2 2 2 0 0 2 0 ...
 $ PEDEMA  : int  0 0 0 2 3 2 0 0 0 2 ...
 $ RESTDYS : int  0 0 2 0 2 2 0 2 2 2 ...
 $ EXERTDYS: int  1 2 3 3 2 3 1 2 3 3 ...
 $ ACTLIMIT: int  1 2 3 3 0 3 0 2 3 2 ...
 $ S3      : int  0 1 0 3 0 2 0 0 2 0 ...
 $ PULCONG : int  2 0 2 2 0 2 0 2 2 0 ...
```

```
$ NSYM    : int  3 3 4 4 4 4 1 4 4 4 ...
$ HEARTRTE: int  96 60 91 85 84 64 102 61 85 112 ...
$ DIABP   : int  80 95 70 80 60 76 90 80 70 80 ...
$ SYSBP   : int  126 142 138 136 120 130 104 138 145 168 ...
$ FUNCTCLS: int  1 3 3 2 1 2 3 1 3 2 ...
$ CHFETIOL: int  1 4 1 2 4 4 1 1 4 1 ...
$ PREVMI  : int  0 1 0 0 0 1 1 0 0 0 ...
$ ANGINA  : int  1 1 1 0 0 0 0 0 1 0 ...
$ DIABETES: int  1 0 0 0 0 0 0 0 0 0 ...
$ HYPERTEN: int  0 1 1 1 0 0 0 1 1 1 ...
$ DIGUSE  : int  1 0 0 1 0 1 1 0 1 1 ...
$ DIURETK : int  0 0 1 0 0 0 0 0 0 0 ...
$ DIURET  : int  1 1 0 1 1 1 1 0 0 1 ...
$ KSUPP   : int  1 0 NA 0 NA 0 NA 0 0 0 ...
$ ACEINHIB: int  1 1 1 1 1 1 1 1 1 1 ...
$ NITRATES: int  0 1 1 1 1 1 1 1 0 0 ...
$ HYDRAL  : int  1 0 0 0 0 0 0 0 0 0 ...
$ VASOD   : int  0 0 0 0 0 0 0 0 0 0 ...
$ DIGDOSE : num  0.25 0.25 0.25 0.25 0.375 0.25 0.25 0.25 0.125 0.25 ...
$ CVD     : int  1 1 1 0 1 1 0 0 0 1 ...
$ CVDDAYS : int  1049 468 631 1157 191 496 903 1369 1747 149 ...
$ WHF     : int  1 1 1 0 1 0 0 0 0 0 ...
$ WHFDAYS : int  1379 1329 631 1157 191 1620 903 1369 1747 1074 ...
$ DIG     : int  0 0 0 0 0 0 0 0 0 0 ...
$ DIGDAYS : int  1438 1360 1391 1157 1550 1620 903 1369 1747 1074 ...
$ MI      : int  0 0 0 0 0 0 0 0 0 0 ...
$ MIDAYS  : int  1438 1360 1391 1157 1550 1620 903 1369 1747 1074 ...
$ UANG    : int  0 0 1 0 0 0 0 0 0 1 ...
$ UANGDAYS: int  1438 1360 746 1157 1550 1620 903 1369 1747 149 ...
$ STRK    : int  0 1 0 0 0 0 0 0 0 0 ...
$ STRKDAYS: int  1438 468 1391 1157 1550 1620 903 1369 1747 1074 ...
$ SVA     : int  0 0 0 0 0 0 0 0 0 0 ...
$ SVADAYS : int  1438 1360 1391 1157 1550 1620 903 1369 1747 1074 ...
$ VENA    : int  0 0 0 0 0 1 0 0 0 0 ...
$ VENADAYS: int  1438 1360 1391 1157 1550 496 903 1369 1747 1074 ...
$ CREV    : int  0 0 0 0 0 0 0 0 0 0 ...
$ CREVDAYS: int  1438 1360 1391 1157 1550 1620 903 1369 1747 1074 ...
$ OCVD    : int  1 0 0 0 0 0 0 0 0 0 ...
$ OCVDDAYS: int  1049 1360 1391 1157 1550 1620 903 1369 1747 1074 ...
$ RINF    : int  0 0 0 0 0 0 0 0 0 0 ...
$ RINFDAYS: int  1438 1360 1391 1157 1550 1620 903 1369 1747 1074 ...
$ OTH     : int  1 1 0 0 1 1 0 0 0 1 ...
$ OTHDAYS : int  533 880 1391 1157 459 966 903 1369 1747 283 ...
$ HOSP    : int  1 1 1 0 1 1 0 0 0 1 ...
$ HOSPDAYS: int  533 468 631 1157 191 496 903 1369 1747 149 ...
$ NHOSP   : int  6 4 2 0 5 5 0 0 0 2 ...
```

```
 $ DEATH   : int  0 1 0 0 0 0 1 0 0 0 ...
 $ DEATHDAY: int  1438 1360 1391 1157 1550 1620 903 1369 1747 1074 ...
 $ REASON  : int  NA 1 NA NA NA NA 2 NA NA NA ...
 $ DWHF    : int  1 1 1 0 1 0 0 0 0 0 ...
 $ DWHFDAYS: int  1379 1329 631 1157 191 1620 903 1369 1747 1074 ...
```

```
> # 6800 rows, 72 columns
```

3. Make a scatterplot with BMI on the X axis and systolic blood pressure on the y axis. Use base graphics

```
> plot(dig$BMI, dig$SYSBP)
```

4. Change the axis titles to "Systolic Blood Pressure" and "Body Mass Index" as appropriate

```
> plot(dig$BMI, dig$SYSBP, xlab="Body Mass Index", ylab="Systolic Blood Pressure")
```

5. Add a title: "BMI and Blood Pressure"

```
> plot(dig$BMI, dig$SYSBP, xlab="Body Mass Index", ylab="Systolic Blood Pressure")
> title("BMI and Blood Pressure")
>
```

6. Make a histogram of ages

```
> hist(dig$AGE)
```

7. Make a histogram of ages such that the title is "Age Histogram"

```
> hist(dig$AGE, main="Age Histogram")
```

8. Create a variable named 'older' that is true for those above the median age

```
> dig$older <- dig$AGE > median(dig$AGE)
```

9. Print out a frequency table of age. How many subjects are in the older group

```
> table(dig$older)

FALSE  TRUE
 3667  3133

> #3133
```

10. Make a boxplot of systolic blood pressure readings by age

```
> boxplot(dig$SYSBP ~ dig$older)
```

11. Change the labels of the boxplot so the left box plot (for younger subjects) is labeled "Younger" and the right is labeled "Older"

```
> boxplot(dig$SYSBP ~ dig$older, names=c("Younger", "Older"))
```

12. Okay, now we'll try ggplot2

13. Make a scatterplot with BMI on the X axis and systolic blood pressure on the y axis using ggplot.

```
> library(ggplot2)
> ggplot(dig) + aes(x=BMI, y=SYSBP) + geom_point()
```

14. Change the axis titles to "Systolic Blood Pressure" and "Body Mass Index" as appropriate

```
> ggplot(dig) +
+   aes(x=BMI, y=SYSBP) +
+   scale_x_continuous(name="Body Mass Index") +
+   scale_y_continuous(name="Systolic Blood Pressure") +
+   geom_point()
```

15. Add a title: "BMI and Blood Pressure"

```
> ggplot(dig) +
+   aes(x=BMI, y=SYSBP) +
+   scale_x_continuous(name="Body Mass Index") +
+   scale_y_continuous(name="Systolic Blood Pressure") +
+   geom_point() +
+   ggtitle("BMI and Blood Pressure")
```

16. Make a histogram of ages

```
> ggplot(dig) +
+   aes(x=AGE) +
+   geom_histogram()
```

17. Add the title is "Age Histogram"

```
> ggplot(dig) +
+   aes(x=AGE) +
+   geom_histogram() +
+   ggtitle("Age Histogram")
```

18. Make a boxplot of systolic blood pressure readings by age

```
> ggplot(dig) +
+   aes(x=older, y=SYSBP) +
+   geom_boxplot()
```

19. Change the labels of the boxplot so the left box plot (for younger subjects) is labeled "Younger" and the right is labeled "Older"

```
> ggplot(dig) +
+   aes(x=older, y=SYSBP) +
+   scale_x_discrete(labels=c("Younger", "Older")) +
+   geom_boxplot()
```