Figures

```
library(tidyverse)
library(readxl)
library(smmr)
```

Figure 1: Packages

> The data file shown in Figure 3 is from a survey of high-school students from different social classes who received high or low parental encouragement to go to college or university, and who said that they did or did not intend to go to college or university. (The last column, `frequency`, is the number of students who fell into that combination of categories.) The data file is called `college-plans.txt`, and is in the folder where you are currently running R.

Figure 2: Scenario A

```
social.stratum;encouragement;college.plans;frequency
lower;low;no;749
lower;low;yes;35
lower;high;no;233
lower;high;yes;133
lowermiddle;low;no;627
lowermiddle;low;yes;38
lowermiddle;high;no;330
lowermiddle;low;no;303
uppermiddle;low;no;627
uppermiddle;low;yes;38
uppermiddle;high;no;374
uppermiddle;high;yes;467
higher;low;no;153
higher;low;yes;26
higher;high;no;266
higher;high;yes;800
```

Figure 3: College plans data set

The file `dogs2.txt`, in the folder where you are currently running R, is shown in Figure 5. The data came from an experiment in which eight dogs were given one of two different drugs, and at times 0, 1, 3, and 5 minutes after the drug was administered, a blood sample was collected and the log of the amount of histamine in the dog's blood was recorded. (A logarithm of an amount can be negative.) Unfortunately, the researcher who collected the data was not very tidy about recording it (although all the values are correct). The text separating the data is spaces (not tabs).

Figure 4: Scenario B

```
Drug    lh0    lh1   lh3     lh5
Morphine      -3.22  -1.61  -2.30  -2.53
Morphine          -3.91  -2.81  -3.91  -3.91
Morphine    -2.66   0.34  -0.73  -1.43
Morphine        -1.77  -0.56 -1.05  -1.43
Trimethaphan        -3.51  -0.48 -1.17  -1.51
Trimethaphan  -3.51   0.05  -0.31  -0.51
Trimethaphan   -2.66  -0.19   0.07 -0.22
Trimethaphan  -2.41   1.14   0.72 0.21
```

Figure 5: Dogs data set

You are working in a lab, and the principal investigator in the lab emails you an Excel spreadsheet called `animals.xlsx` (as an attachment to the email, which you can read but not edit, or save in any other form), containing some animal data that you need to analyze, in `Sheet1`.

Figure 6: Scenario C

```
# A tibble: 20 x 6
      rw   fpg glucose insulin  sspg group
   <dbl> <int>   <int>   <int> <int> <chr>
 1  1.04   203     967     138   351 overt
 2  0.97    86     393     115    85 normal
 3  0.91   100     350     221   119 normal
 4  1.07   104     472     180   239 chemical
 5  0.78    98     321     222    99 normal
 6  1.2     89     472     162   257 chemical
 7  0.99    97     379     142    98 normal
 8  1.2    102     472     297   272 chemical
 9  1.18    96     418     130   153 normal
10  1.16   112     562     139   198 chemical
11  0.76    86     381     157   165 normal
12  0.9    213    1025      29   209 overt
13  1       99     336     143   105 normal
14  1.05    96     456     326   235 chemical
15  1.06   151     854      76   260 overt
16  0.85   216    1113      81   378 overt
17  0.92   303    1364      42   346 overt
18  0.95    96     356     112    73 normal
19  1.05   110     477     124    60 chemical
20  0.76    90     353     263   165 normal
```

Figure 7: Diabetes data (20 randomly chosen rows out of 145 observations)

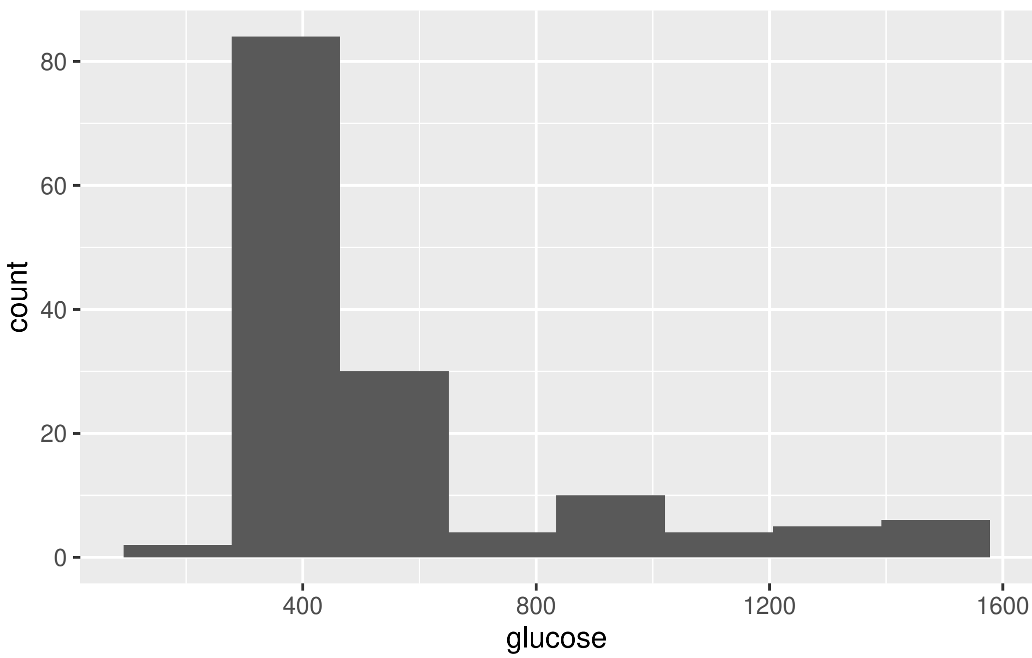

Figure 8: Graph of diabetes data

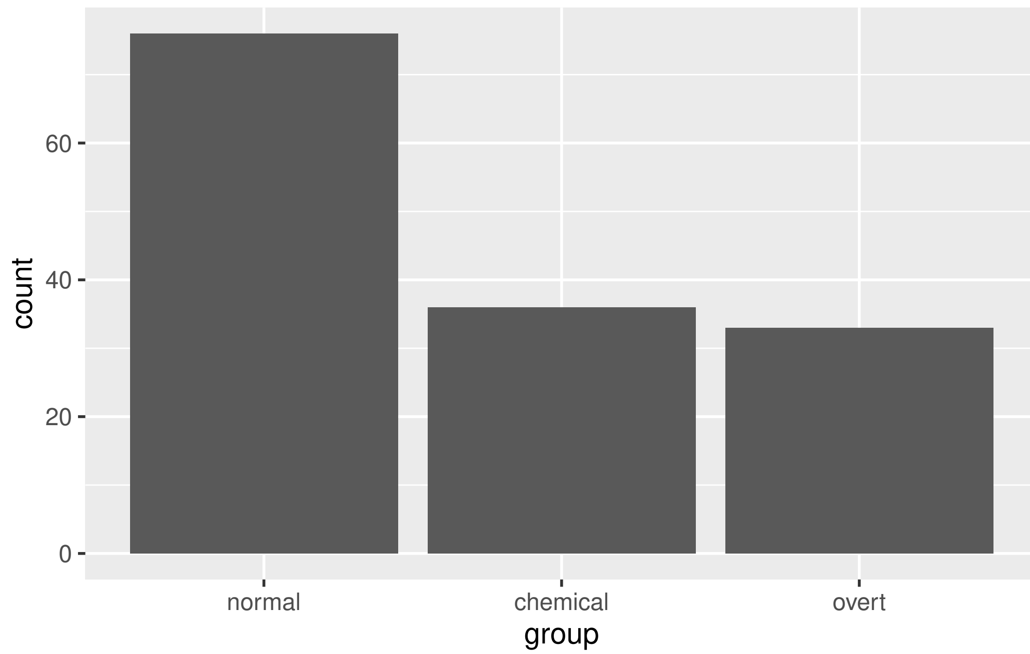

Figure 9: Another graph of diabetes data

```
# A tibble: 100 x 1
   heights
     <int>
 1      71
 2      67
 3      69
 4      70
 5      68
 6      63
 7      68
 8      72
 9      70
10      70
# i 90 more rows
```

Figure 10: Men's heights (first 10 rows)

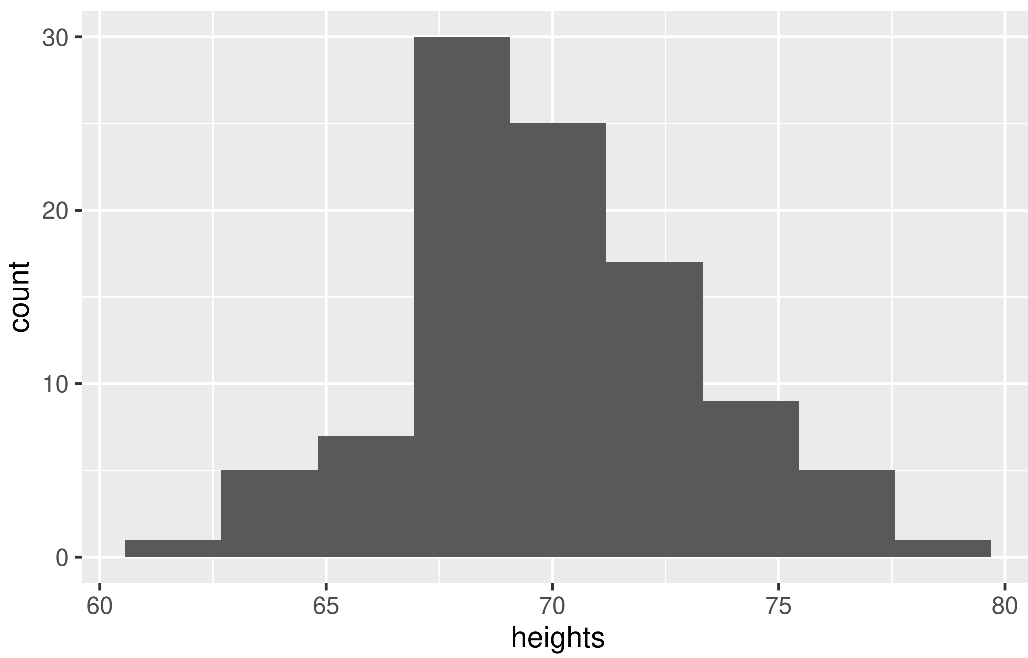

Figure 11: Histogram of heights

```
    One Sample t-test

data:  heights
t = 2.825, df = 99, p-value = 0.005719
alternative hypothesis: true mean is not equal to 69.1
95 percent confidence interval:
 69.37679 70.68321
sample estimates:
mean of x
    70.03
```

Figure 12: Male heights t-test

```
tibble(sim = 1:10000) %>%
  rowwise() %>%
  mutate(my_sample = list(sample(male_heights$heights, replace = TRUE))) %>%
  mutate(my_mean = mean(my_sample)) %>%
  ggplot(aes(sample = my_mean)) + stat_qq() + stat_qq_line()
```

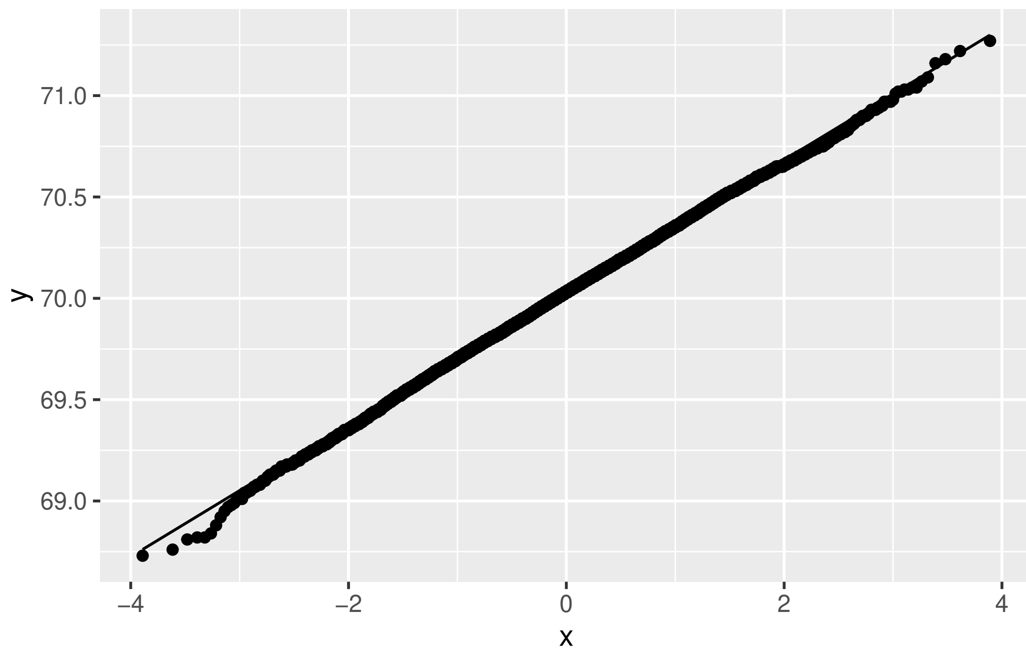

Figure 13: Code and output for heights data

```
# A tibble: 20 x 2
   Condition   Rom
   <chr>     <dbl>
 1 LBP        88.7
 2 No LBP     95.4
 3 No LBP     92.7
 4 No LBP     91.9
 5 No LBP     96.7
 6 LBP        83.3
 7 LBP        72.3
 8 No LBP     94.1
 9 LBP        91.3
10 No LBP     82.4
11 No LBP     90.0
12 No LBP     85.6
13 LBP        89.9
14 No LBP    100.
15 No LBP     97.5
16 No LBP     82.0
17 No LBP     98.5
18 LBP        79.1
19 No LBP     95.3
20 LBP        88.4
```

Figure 14: Lower back pain data (in dataframe `lbp`), 20 randomly chosen rows

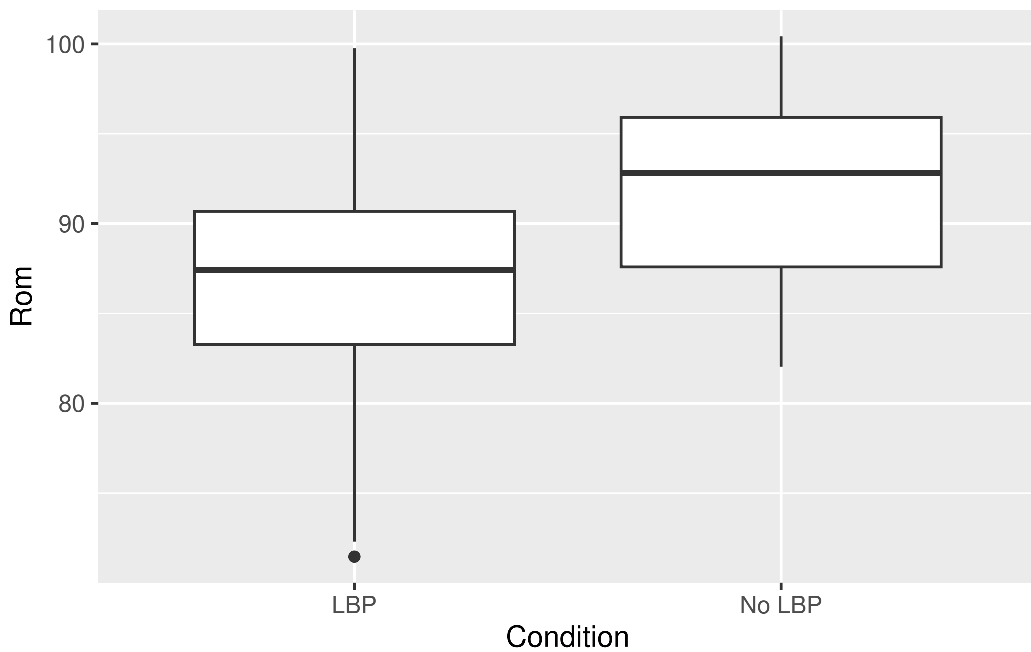

Figure 15: Boxplot of lower back pain data

```
tibble(sim = 1:1000) %>%
  rowwise() %>%
  mutate(my_sample = list(sample(rnorm(50, 40, 15)))) %>%
  mutate(t_test = list(t.test(my_sample, mu = 45))) %>%
  mutate(p_value = t_test$p.value) %>%
  count(p_value <= 0.05)
```

```
# A tibble: 2 x 2
  `p_value <= 0.05`     n
  <lgl>             <int>
1 FALSE               380
2 TRUE                620
```

Figure 16: Mystery code

```
     Two-sample t test power calculation

              n = 9.283698
          delta = 0.8
             sd = 0.5
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number in *each* group
```

Figure 17: Output from your previous code

```
# A tibble: 20 x 1
   score
   <int>
 1    40
 2     4
 3    19
 4     8
 5    28
 6    25
 7    19
 8    19
 9    26
10    18
11    78
12    11
13     8
14    65
15    17
16    26
17    11
18    19
19     7
20    94
```

Figure 18: Wisconsin card-sorting test data (20 randomly chosen rows) in dataframe `wcst`

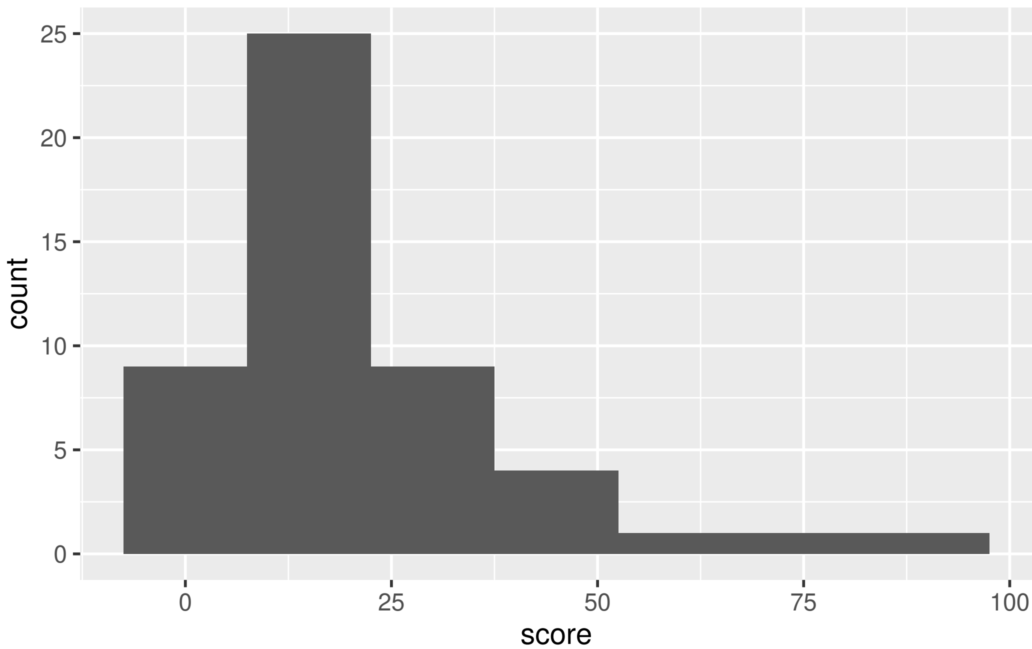

Figure 19: Histogram of Wisconsin card-sorting test data

```
[1] 11.00488 22.99854
```

Figure 20: Confidence interval for median `score`

```
# A tibble: 20 x 2
      m1 lecture
   <dbl> <fct>
 1    89 a
 2    86 b
 3    79 a
 4    83 b
 5    95 c
 6    63 a
 7    48 b
 8    59 c
 9    85 b
10    71 a
11    80 a
12    70 a
13    99 a
14    63 a
15    62 b
16    50 a
17    73 b
18    97 b
19   100 a
20    81 a
```

Figure 21: Lecture section data (20 randomly chosen rows)

```
             Df Sum Sq Mean Sq F value Pr(>F)
lecture       2   1290   645.1   3.484  0.033 *
Residuals   161  29810   185.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 22: Lecture section analysis 1

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = m1 ~ lecture, data = students)

$lecture
         diff        lwr       upr     p adj
b-a -3.139812 -9.1981360  2.918512 0.4395988
c-a  3.837728 -2.3413371 10.016793 0.3084418
c-b  6.977540  0.7201228 13.234957 0.0247211
```

Figure 23: Lecture section analysis 2