Figures

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

Figure 1: Packages

time_of_day	download_time
early	69
early	138
early	75
early	186
early	68
early	217
evening	299
evening	367
evening	331
evening	257
evening	260
evening	269
late-night	216
late-night	175
late-night	274
late-night	171
late-night	187
late-night	213

Figure 2: File download data (selected rows out of 48 total rows)

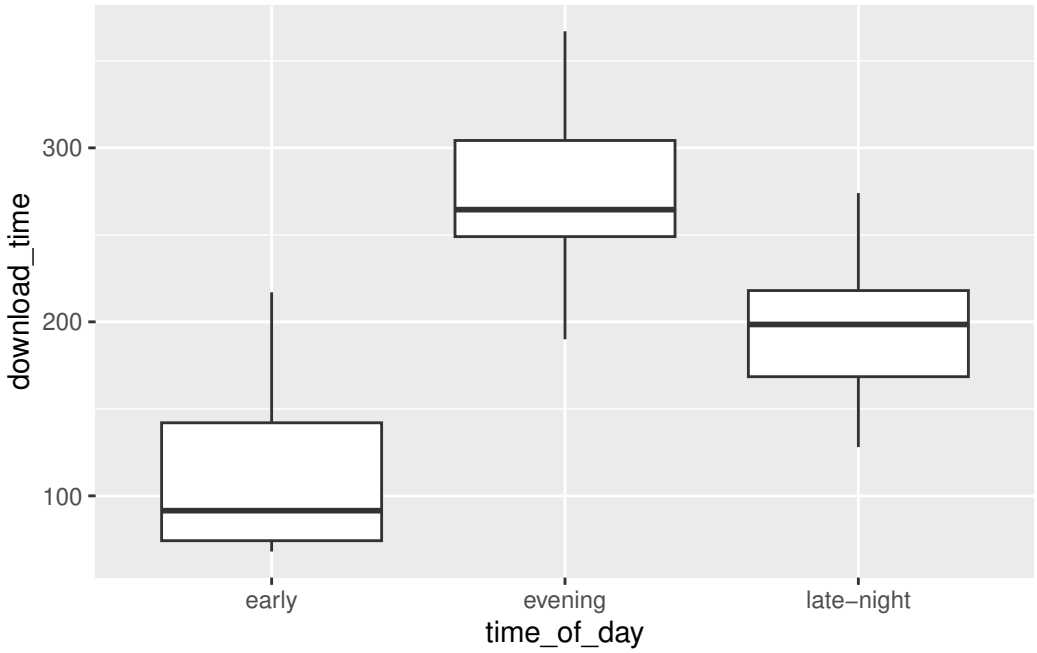

Figure 3: Boxplot for downloading data

```
# A tibble: 47 x 3
    Year Gender Height
   <dbl> <chr>   <dbl>
 1  1896 Men      1.81
 2  1900 Men      1.9
 3  1904 Men      1.8
 4  1908 Men      1.90
 5  1912 Men      1.93
 6  1920 Men      1.94
 7  1924 Men      1.98
 8  1928 Men      1.94
 9  1932 Men      1.97
10  1936 Men      2.03
# i 37 more rows
```

Figure 4: The high-jump data

```
subject hypnotized score
1 yes 8.5
2 yes 9.6
3 yes 10.0
4 yes 9.2
5 yes 8.9
6 yes 10.8
7 no 12.6
8 no 13.8
9 no 11.6
10 no 12.2
11 no 12.1
12 no 13.0
```

Figure 5: The Stroop test data

```
    Two Sample t-test

data:  score by hypnotized
t = 6.606, df = 10, p-value = 3.016e-05
alternative hypothesis: true difference in means between group no and group yes is greater
95 percent confidence interval:
 2.213187      Inf
sample estimates:
 mean in group no mean in group yes
            12.55              9.50
```

Figure 6: Stroop data t-test

```
# A tibble: 202 x 13
   Sex    Sport     RCC   WCC    Hc    Hg  Ferr   BMI   SSF `%Bfat`   LBM    Ht
   <chr>  <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl>
 1 female Netball  4.56  13.3  42.2  13.6    20  19.2  49      11.3  53.1  177.
 2 female Netball  4.15   6    38    12.7    59  21.2 110.     25.3  47.1  173.
 3 female Netball  4.16   7.6  37.5  12.3    22  21.4  89      19.4  53.4  176
 4 female Netball  4.32   6.4  37.7  12.3    30  21.0  98.3    19.6  48.8  170.
 5 female Netball  4.06   5.8  38.7  12.8    78  21.8 122.     23.1  56.0  183
 6 female Netball  4.12   6.1  36.6  11.8    21  21.4  90.4    16.9  56.4  178.
 7 female Netball  4.17   5    37.4  12.7   109  21.5 107.     21.3  53.1  177.
 8 female Netball  3.8    6.6  36.5  12.4   102  24.4 157.     26.6  54.4  174.
 9 female Netball  3.96   5.5  36.3  12.4    71  22.6 101.     17.9  56.0  174.
10 female Netball  4.44   9.7  41.4  14.1    64  22.8 126.     25.0  51.6  174.
11 female Netball  4.27  10.6  37.7  12.5    68  23.6 114      22.6  58.3  179.
12 female Netball  3.9    6.3  35.9  12.1    78  20.1  70      15.0  57.3  183.
13 female Netball  4.02   9.1  37.7  12.7   107  23.0  77      18.1  57.3  174.
14 female Netball  4.39   9.6  38.3  12.5    39  24.6 149.     26.8  54.2  173.
15 female Netball  4.52   5.1  38.8  13.1    58  18.3  80.1    17.2  43.0  169.
16 female Netball  4.25  10.7  39.5  13.2   127  24.5 157.     26.5  54.5  174
17 female Netball  4.46  10.9  39.7  13.7   102  24.0 116.     23.0  57.2  176
18 female Netball  4.4    9.3  40.4  13.6    86  26.2 182.     30.1  54.4  172.
19 female Netball  4.83   8.4  41.8  13.4    40  20.0  71.6    13.9  57.6  183.
20 female Netball  4.23   6.9  38.3  12.6    50  25.7 144.     26.6  61.5  180.
21 female Netball  4.24   8.4  37.6  12.5    58  25.6 201.     35.5  53.5  180.
22 female Netball  3.95   6.6  38.4  12.8    33  19.9  68.9    15.6  54.1  180.
23 female Netball  4.03   8.5  37.7  13      51  23.4 104.     19.6  55.4  172.
24 female BBall    3.96   7.5  37.5  12.3    60  20.6 109.     19.8  63.3  196.
25 female BBall    4.41   8.3  38.2  12.7    68  20.7 103.     21.3  58.6  190.
26 female BBall    4.14   5    36.4  11.6    21  21.9 105.     19.9  55.4  178.
27 female BBall    4.11   5.3  37.3  12.6    69  21.9 126.     23.7  57.2  185
28 female BBall    4.45   6.8  41.5  14      29  19.0  80.3    17.6  53.2  185.
29 female BBall    4.1    4.4  37.4  12.5    42  21.0  75.2    15.6  53.8  174
30 female BBall    4.31   5.3  39.6  12.8    73  21.7  87.2    20.0  60.2  186.
# i 172 more rows
# i 1 more variable: Wt <dbl>
```

Figure 7: Australian athletes data (some)

```
ggplot(athletes,aes(x=BMI))+geom_histogram(bins=8)
```

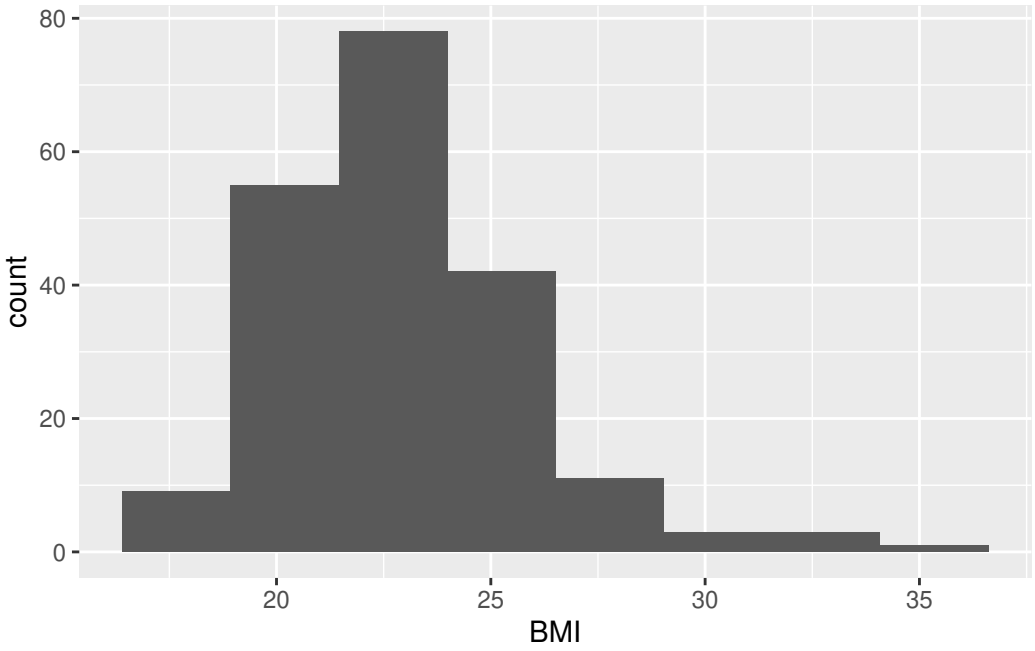

Figure 8: Australian athletes BMI histogram

Note that `t.test` can also do a (Welch) two-sample t-test by inputting the observations for the two groups in separate vectors.

```
tibble(sim = 1:1000) %>%
  rowwise() %>%
  mutate(sample1 = list(rnorm(25, 60, 8))) %>%
  mutate(sample2 = list(rnorm(22, 55, 10))) %>%
  mutate(t_test = list(t.test(sample1, sample2))) %>%
  mutate(p_val = t_test$p.value) %>%
  count(p_val <=0.05)
```

```
# A tibble: 2 x 2
  `p_val <= 0.05`     n
  <lgl>           <int>
1 FALSE             544
2 TRUE              456
```

Figure 9: Some R code

```
# A tibble: 33 x 1
     age
   <dbl>
 1  35.5
 2  44.5
 3  39.8
 4  33.3
 5  51.4
 6  51.3
 7  30.5
 8  48.9
 9  42.1
10  40.3
11  46.8
12  38
13  40.1
14  36.8
15  39.3
16  71.1
17  73.4
18  65.4
19  42.6
20  42.8
21  59.8
22  52.4
23  26.2
24  60.9
25  45.6
26  27.1
27  47.3
28  36.6
29  23.2
30  55.6
31  45.1
32  52.2
33  43.5
```

Figure 10: Diabetes patients data

```
ggplot(diabetes, aes(x = age)) + geom_histogram(bins = 7)
```

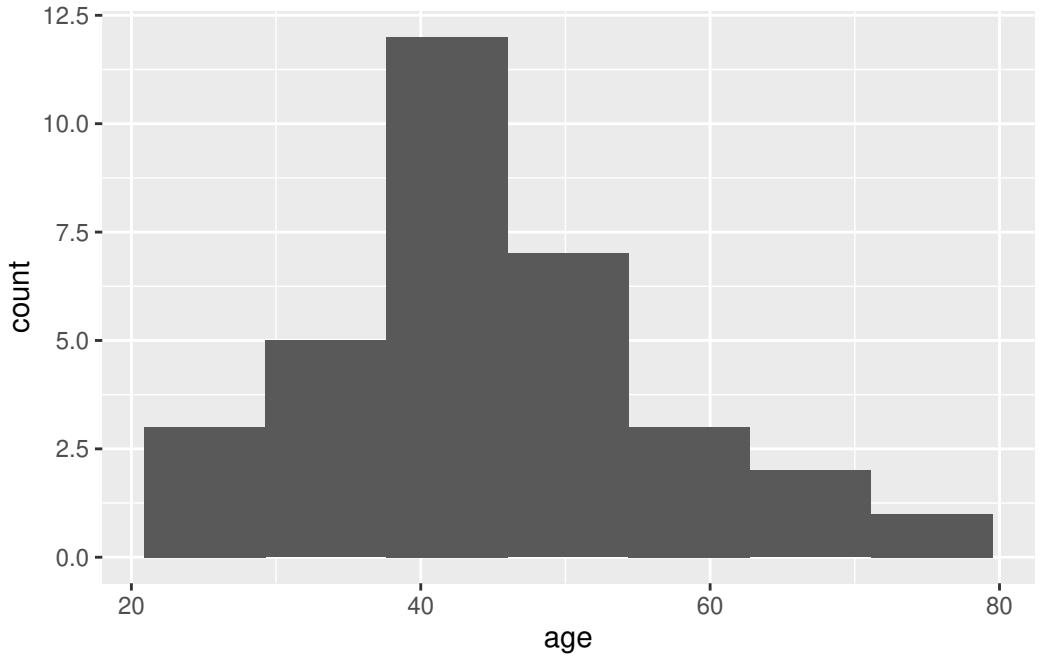

Figure 11: Histogram of diabetes data

```
sign_test(diabetes, age, 37)
```

```
$above_below
below above
    8    25

$p_values
  alternative     p_value
1       lower 0.999340636
2       upper 0.002275692
3   two-sided 0.004551384
```

```
proc univariate location=37;
  var age;
```

Figure 12: Diabetes tests for location

```
# xxx
diabetes = read_delim("diabetes.txt"," ")
```

```
Rows: 33 Columns: 1
-- Column specification --------------------------------------------------------
Delimiter: " "
dbl (1): age

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
diabetes %>% count(age<37)
```

```
# A tibble: 2 x 2
  `age < 37`     n
  <lgl>      <int>
1 FALSE         25
2 TRUE           8
```

```
succ=20:33
tibble(succ,prob=dbinom(succ,33,0.5)) %>% print(n=Inf)
```

```
# A tibble: 14 x 2
    succ     prob
   <int>    <dbl>
 1    20 6.67e- 2
 2    21 4.13e- 2
 3    22 2.25e- 2
 4    23 1.08e- 2
 5    24 4.49e- 3
 6    25 1.62e- 3
 7    26 4.97e- 4
 8    27 1.29e- 4
 9    28 2.76e- 5
10    29 4.76e- 6
11    30 6.35e- 7
12    31 6.15e- 8
13    32 3.84e- 9
14    33 1.16e-10
```

Figure 13: R output for diabetes data analysis

```
Rows: 10 Columns: 2
-- Column specification --------------------------------------------------------
Delimiter: "\t"
dbl (2): bottom, surface

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

# A tibble: 10 x 2
   bottom surface
    <dbl>   <dbl>
 1  0.43    0.415
 2  0.266   0.238
 3  0.567   0.39
 4  0.531   0.41
 5  0.707   0.605
 6  0.716   0.609
 7  0.651   0.632
 8  0.589   0.523
 9  0.469   0.411
10  0.723   0.612
```

Figure 14: Zinc concentration data

```
with(zinc,t.test(surface, bottom, alternative="greater"))
```

```
    Welch Two Sample t-test

data:  surface and bottom
t = -1.2913, df = 17.779, p-value = 0.8934
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.1884396        Inf
sample estimates:
mean of x mean of y
   0.4845    0.5649
```

Figure 15: Zinc t-test 1

```
with(zinc, t.test(bottom, surface, paired=TRUE, alternative="greater"))
```

```
    Paired t-test

data:  bottom and surface
t = 4.8638, df = 9, p-value = 0.0004456
alternative hypothesis: true mean difference is greater than 0
95 percent confidence interval:
 0.0500982       Inf
sample estimates:
mean difference
         0.0804
```

Figure 16: Zinc t-test 2

```
ggplot(zinc, aes(sample = surface)) + stat_qq() + stat_qq_line()
```

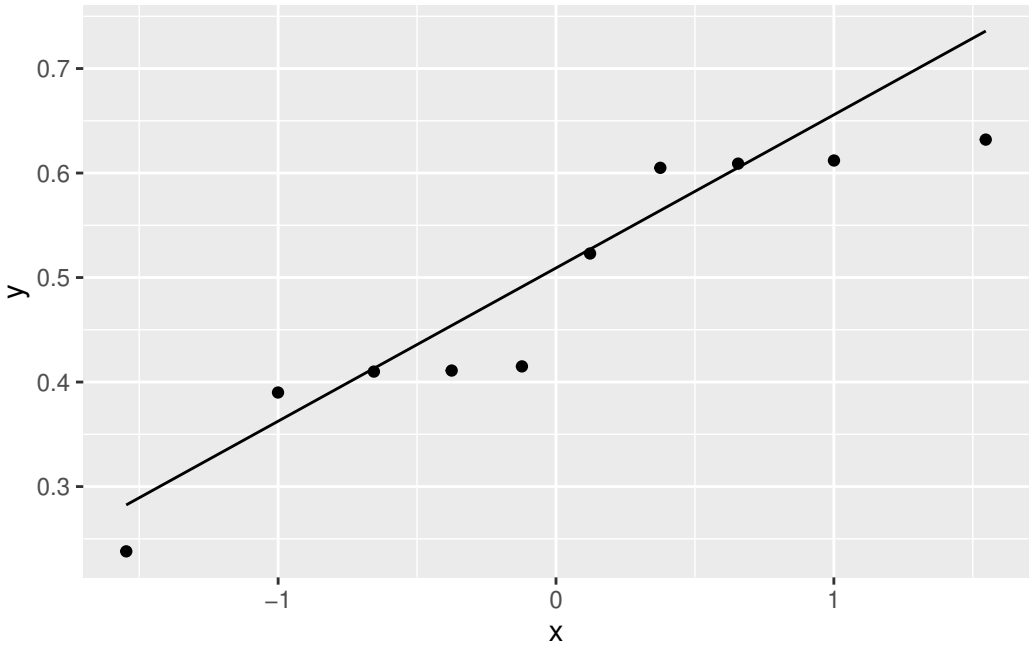

Figure 17: Zinc normal quantile plot of surface measurements

```
ggplot(zinc, aes(sample = bottom)) + stat_qq() + stat_qq_line()
```

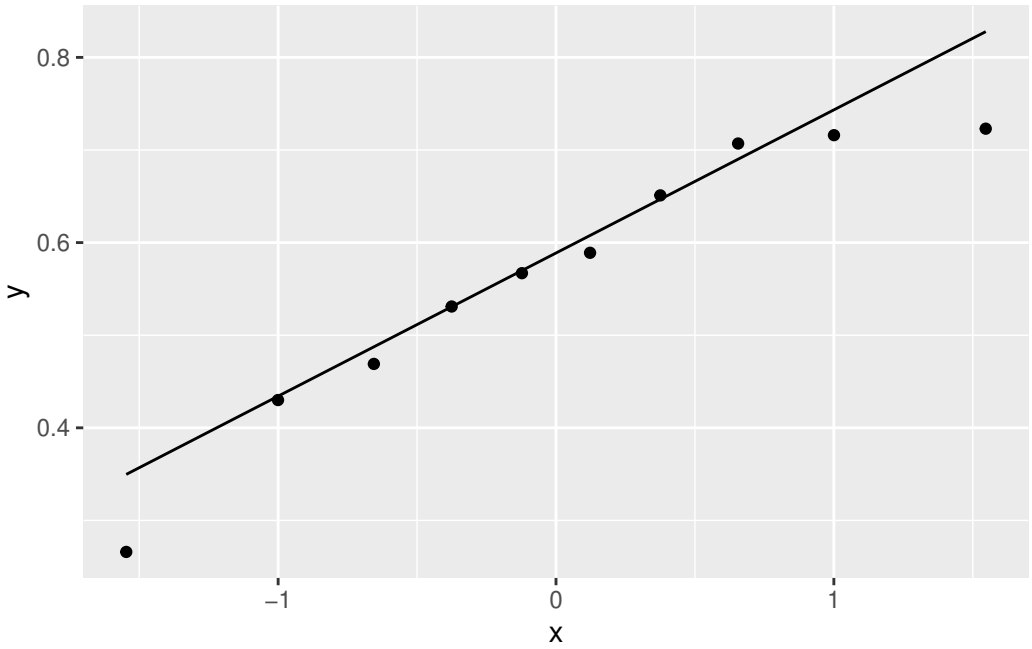

Figure 18: Zinc normal quantile plot of bottom measurements

```
zinc %>% mutate(diff = bottom - surface) %>%
  ggplot(aes(sample = diff)) + stat_qq() + stat_qq_line()
```

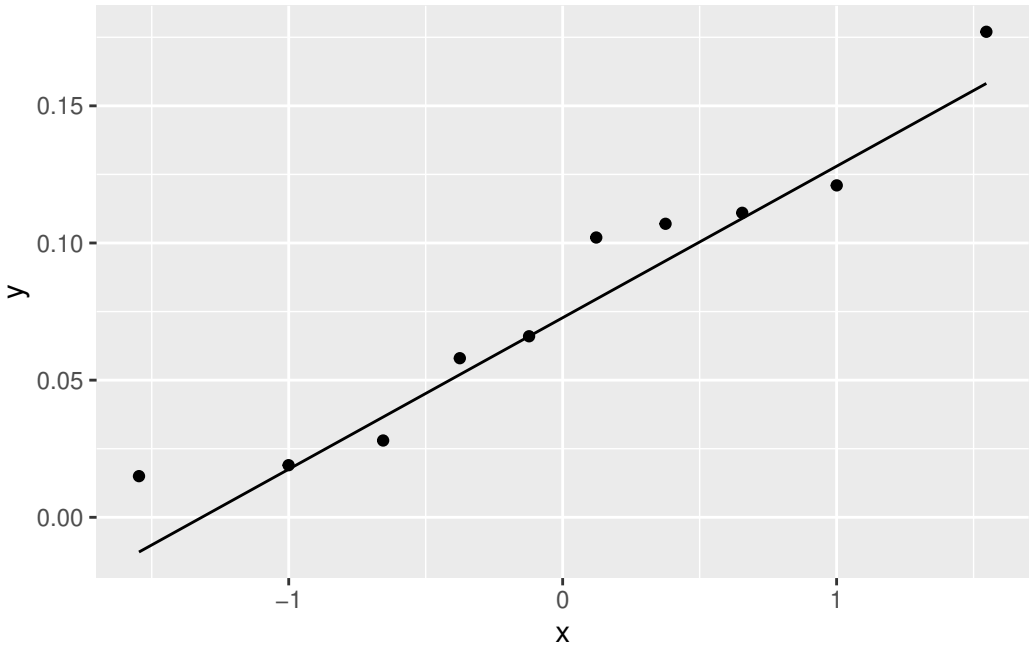

Figure 19: Zinc normal quantile plot of differences

```
prenatal <- read_csv("prenatal.csv")
```

```
Rows: 15 Columns: 2
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): care
dbl (1): apgar

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
prenatal
```

```
# A tibble: 15 x 2
   care   apgar
   <chr>  <dbl>
 1 usual      8
 2 usual      7
 3 usual      6
 4 usual      5
 5 usual      2
 6 usual      8
 7 usual      7
 8 usual      3
 9 visits     9
10 visits     9
11 visits     7
12 visits     8
13 visits    10
14 visits     9
15 visits     6
```

Figure 20: Prenatal care data

```
library(smmr)
median_test(prenatal,apgar,care)
```

```
$grand_median
[1] 7

$table
        above
group    above below
  usual      2     4
  visits     5     1

$test
       what      value
1 statistic 3.08571429
2        df 1.00000000
3   P-value 0.07898258
```

Figure 21: Hypothesis test for prenatal care data