Baseball Assignment

Katie, Rita, and Chang

2023-09-19

Problem 1 - nHits

Subgroups

```
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
infield <- filter(dat, Position == '1B' | Position == '2B' | Position == 'SS' | Position
== '3B')
outfield <- filter(dat, Position == 'CF' | Position == 'RF' | Position == 'LF' | Positio
n == 'OF')
catcher <- filter(dat, Position == 'C')
CrHits2 <- dat$CrHits*dat$CrHits
```

Plot League by Division

```
library("arsenal")
tab.lbyd <- table(dat$League, dat$Division)
lbyd <- freqlist(tab.lbyd)
summary(lbyd)
```

```
##
##
## |Var1     |Var2 | Freq| Cumulative Freq| Percent| Cumulative Percent|
## |:--------|:----|----:|---------------:|-------:|------------------:|
## |American |East |   85|              85|   26.40|              26.40|
## |         |West |   90|             175|   27.95|              54.35|
## |National |East |   72|             247|   22.36|              76.71|
## |         |West |   75|             322|   23.29|             100.00|
```

Plot Division for American Leagues

```
library("arsenal")
library("dplyr")
data <- filter(dat, League == 'American')
tab.adiv <- table(data$Division)
adiv <- freqlist(tab.adiv)
summary(adiv)
```

```
##
##
## |Var1 | Freq| Cumulative Freq| Percent| Cumulative Percent|
## |:----|----:|---------------:|-------:|------------------:|
## |East |   85|              85|   48.57|              48.57|
## |West |   90|             175|   51.43|             100.00|
```

Arrange by Division

```
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns))
print(data[order(data$Division),])
```

```
##        Salary        Team Division YrMajor logSalary nAtBat nBB nError nHits
## 1          NA   Cleveland     East       1        NA    293  14     20    66
## 5    1100.000   Cleveland     East      13  7.003065    401  65      0    92
## 6     517.143     Detroit     East      10  6.248319    574  59     22   159
## 7     700.000   Baltimore     East       6  6.551080    239  22      6    60
## 13    776.667      Boston     East      18  6.655012    629  40     14   168
## 14    765.000   Cleveland     East       6  6.639876    587  70      3   163
## 17    612.500   Cleveland     East       5  6.417549    583  56     25   168
## 19         NA    New York     East       4        NA    161  17     12    36
## 20         NA   Milwaukee     East      16        NA    346  30      0    98
## 21     67.500   Milwaukee     East       2  4.212128    181  33      5    41
## 22    180.000   Milwaukee     East       4  5.192957    217   9      1    46
## 23         NA    New York     East      11        NA    194  30      2    40
## 24    305.000   Cleveland     East       6  5.720312    254  22      4    68
## 25    247.500   Cleveland     East       5  5.511411    205   9      4    57
## 26         NA   Milwaukee     East      16        NA    542  41      9   140
## 28         NA     Toronto     East      15        NA    336  52      0    84
## 30    675.000     Detroit     East      12  6.514713    403  39      5   101
## 31         NA   Milwaukee     East      14        NA    235  21      4    61
## 32   1350.000   Baltimore     East       6  7.207860    627  70     13   177
## 33     90.000   Cleveland     East       1  4.499810    416  16     10   113
## 36    950.000      Boston     East      17  6.856462    585  62      0   139
## 38    105.000     Detroit     East       4  4.653960    521  45     23   142
## 39         NA     Detroit     East      12        NA    419  44      1   113
## 41    535.000     Detroit     East      18  6.282267    507  91      2   122
## 42    933.333      Boston     East      15  6.838762    529  97      5   137
## 43    850.000     Toronto     East       9  6.745236    424  13      8   119
## 47   1975.000    New York     East       5  7.588324    677  53      6   238
## 50    110.000    New York     East       2  4.700480    280  47      2    82
## 54     70.000   Milwaukee     East       1  4.248495    317  32     26    78
## 56   1861.460    New York     East      14  7.529116    565  77      5   148
## 57   2460.000   Baltimore     East      10  7.807917    495  78     13   151
## 58         NA   Milwaukee     East       2        NA    524  54     20   132
## 59    375.000      Boston     East       8  5.926926    233  18     10    49
## 60         NA     Toronto     East      10        NA    395  35      7   106
## 61         NA   Baltimore     East      13        NA    397  53      4   114
## 62         NA   Baltimore     East       6        NA    210  15     15    37
## 64   1175.000     Toronto     East       5  7.069023    641  41     10   198
## 65     70.000   Milwaukee     East       1  4.248495    215  11     12    51
## 70    362.500     Toronto     East       8  5.893024    327  20     12    85
## 78   1237.500     Toronto     East       6  7.120848    589  69      3   170
## 79    430.000   Baltimore     East      15  6.063785    343  40     13   103
## 80         NA   Baltimore     East       5        NA    284  25      5    69
## 83    250.000   Cleveland     East       4  5.521461    663  32      6   200
## 85    275.000   Baltimore     East      14  5.616771    160  22      0    39
## 86    775.000   Cleveland     East       5  6.652863    599  32     18   183
## 87    850.000   Milwaukee     East      11  6.745236    497  26     10   136
## 88    365.000     Detroit     East      15  5.899897    210  28      0    70
## 96   2412.500      Boston     East      13  7.788419    618  62      8   200
## 97    300.000   Baltimore     East       6  5.703782    404  18      5    92
## 100        NA   Baltimore     East       2        NA    212  18      5    54
## 102  1300.000     Detroit     East       8  7.170120    441  68      2   118
```

```
## 103 1000.000    New York     East        14  6.907755    490  35       3   150
## 107  225.000     Detroit     East        12  5.416100    283  27       2    70
## 108  525.000   Baltimore     East        15  6.263398    491  37       2   141
## 109  787.500     Toronto     East         7  6.668863    589  64       6   149
## 110  800.000     Detroit     East        10  6.684612    327  38       6    84
## 112  145.000   Baltimore     East         3  4.976734    338  21       0    92
## 114  420.000     Detroit     East        10  6.040255    584  63      11   157
## 115  575.000      Boston     East         5  6.354370    625  65      14   179
## 117  700.000    New York     East        13  6.551080    490  49       0   148
## 118  550.000   Cleveland     East         6  6.309918    442  33       7   131
## 121  175.000    New York     East         3  5.164786    504  54      19   120
## 124  350.000   Baltimore     East         5  5.857933    369  49       6    93
## 129 1260.000   Milwaukee     East         9  7.138867    437  40      15   123
## 131  190.000     Detroit     East         5  5.247024    236  21       4    56
## 132  580.000   Cleveland     East         6  6.363028    473  29       9   154
## 135  250.000   Milwaukee     East        12  5.521461    216  15       4    56
## 136  215.000   Milwaukee     East         3  5.370638    466  72       8   108
## 137  400.000   Baltimore     East        18  5.991465    327  45       7    68
## 138       NA      Boston     East         7        NA    462  37       6   119
## 139  560.000    New York     East         9  6.327937    341  46       4   110
## 140 1670.000    New York     East         8  7.420579    608  89       6   160
## 145  250.000     Toronto     East         6  5.521461    246  13       1    76
## 146  400.000   Milwaukee     East        12  5.991465    205  17       2    52
## 147  450.000     Toronto     East        10  6.109248    348  43       6    90
## 148   70.000      Boston     East         1  4.248495    312  24      15    68
## 152 1000.000   Milwaukee     East        13  6.907755    522  62       1   163
## 159       NA      Boston     East        11        NA    425  24       8   112
## 160  530.000   Cleveland     East         8  6.272877    562  53      17   169
## 161  341.667     Detroit     East         8  5.833837    281  20       7    76
## 163  350.000     Toronto     East         4  5.857933    687  27      13   213
## 167       NA   Baltimore     East         5        NA    181  17       9    46
## 170 1600.000      Boston     East         5  7.377759    580 105      19   207
## 172  875.000    New York     East        12  6.774224    492  94      20   136
## 174  960.000     Toronto     East         8  6.866933    573  78      12   144
## 2    480.000     Seattle     West         3  6.173786    479  76      14   130
## 3    750.000     Oakland     West        11  6.620073    594  35      25   169
## 4    100.000 Kansas City     West         3  4.605170    298   7       9    73
## 8         NA Minneapolis     West         3        NA    183  11       0    39
## 9    175.000 Kansas City     West         5  5.164786    190  15      16    46
## 10        NA     Oakland     West        12        NA    407  65       9   104
## 11   115.000     Chicago     West         1  4.744932    426  62       2   109
## 12        NA  California     West        15        NA    442  43      11    98
## 15   900.000  California     West        14  6.802395    513  90       3   137
## 16        NA  California     West        17        NA    313  39       7    84
## 18   300.000     Seattle     West         7  5.703782    204  12       5    49
## 27   875.000     Chicago     West        17  6.774224    457  22       4   101
## 29  1200.000     Oakland     West         9  7.090077    591  39       4   168
## 34   230.000       Texas     West         4  5.438079    236  11      13    56
## 35        NA     Oakland     West        19        NA    242  27       0    58
## 37    75.000     Chicago     West         3  4.317488    199  21       5    53
## 40   850.000  California     West        14  6.745236    512  52      12   131
## 44   325.000     Seattle     West         6  5.783825    388  39       4   103
```

```
## 45   275.000     Oakland     West       4  5.616771    339  23      9   96
## 46        NA     Oakland     West      16        NA    561  33      8  118
## 48        NA Kansas City     West       5        NA    227  12      2   46
## 49   600.000     Oakland     West       9  6.396930    329  56      2   83
## 51   260.000       Texas     West      16  5.560682    155  22      1   41
## 52   475.000  California     West       4  6.163315    458  48     18  114
## 53   431.500       Texas     West       5  6.067268    314  16      4   83
## 55   145.000     Seattle     West       3  4.976734    511  61      8  138
## 63   750.000 Kansas City     West      14  6.620073    566  43     10  154
## 66  1500.000 Kansas City     West      14  7.313220    441  80     16  128
## 67   900.000 Minneapolis     West       6  6.802395    596  52     21  171
## 68   155.000 Minneapolis     West       4  5.043425    472  30     26  118
## 69   700.000  California     West      16  6.551080    283  26      5   77
## 71   400.000  California     West       5  5.991465    539  69      7  139
## 72        NA     Seattle     West      13        NA    315  58      0   59
## 73   500.000     Chicago     West       5  6.214608    282  29      5   78
## 74   600.000       Texas     West       8  6.396930    380  31      1  120
## 75   950.000     Chicago     West       7  6.856462    570  38      5  169
## 76   325.000 Kansas City     West      18  5.783825    278  18      0   70
## 77    87.500     Seattle     West       4  4.471639    445  29     16   99
## 81   100.000     Chicago     West       2  4.605170    438  71      9  103
## 82   165.000     Oakland     West       2  5.105945    600  65     14  144
## 84        NA     Chicago     West      10        NA    209  42      5   45
## 89        NA     Chicago     West      11        NA    225  26      0   61
## 90    95.000  California     West       2  4.553877    151  19      2   41
## 91    80.000     Seattle     West       5  4.382027    399  34      3  102
## 92        NA Kansas City     West      15        NA    336  23      0   93
## 93   200.000     Seattle     West       3  5.298317    616  32     15  163
## 94        NA Kansas City     West      12        NA    219  17      4   47
## 95    75.000 Minneapolis     West       3  4.317488    165  16      2   39
## 98   110.000     Chicago     West       4  4.700480    315  16      3   73
## 99   825.000 Kansas City     West      13  6.715383    429  57      4   91
## 101       NA     Oakland     West       3        NA    161  22      2   43
## 104 1310.000 Minneapolis     West       6  7.177782    550  71     10  147
## 105  300.000     Seattle     West       7  5.703782    344  88      0   85
## 106  365.000 Minneapolis     West       3  5.899897    680  34      6  223
## 111  587.500       Texas     West      13  6.375876    464  52      0  128
## 113       NA Kansas City     West       9        NA    508  46      9  146
## 116  780.000     Oakland     West       7  6.659294    489  34      9  131
## 119       NA Minneapolis     West       8        NA    317  19      4   88
## 120   68.000 Kansas City     West       1  4.219508    209  12      3   54
## 122  137.000 Minneapolis     West       3  4.919981    258  18      8   60
## 123  120.000     Oakland     West       3  4.787492    211  39      8   43
## 125  175.000     Chicago     West       2  5.164786    547  12     22  137
## 126  200.000       Texas     West       2  5.298317    572  65      3  152
## 127  750.000     Seattle     West       4  6.620073    526  77      1  163
## 128  172.000       Texas     West       1  5.147494    540  55     14  135
## 130       NA       Texas     West       5        NA    551  87     11  160
## 133  450.000  California     West      12  6.109248    271  33      3   77
## 134  300.000 Minneapolis     West       5  5.703782    357  39      4   96
## 141  487.500  California     West      20  6.189290    419  92      0  101
## 142       NA  California     West      11        NA    393  64      4   90
```

```
## 143  425.000     Chicago     West         5  6.052089    376  35       0    82
## 144       NA Kansas City     West         7        NA    307  29       2    80
## 149   97.500       Texas     West         1  4.579852    382  22       6   101
## 150  740.000 Minneapolis     West        12  6.606650    459  68       0   113
## 151  341.667  California     West        10  5.833837    288  16       7    63
## 153  100.000 Kansas City     West         6  4.605170    512  43      18   117
## 154   90.000     Seattle     West         3  4.499810    220  13       3    66
## 155  135.000       Texas     West         2  4.905275    461  35      11   112
## 156  475.000       Texas     West         6  6.163315    530  47      15   159
## 157  105.000 Minneapolis     West         2  4.653960    453  52       6   103
## 158  350.000     Seattle     West         4  5.857933    528  51      17   122
## 162  940.000 Minneapolis     West         6  6.845880    593  53       6   152
## 164       NA       Texas     West        17        NA    289  44       7    63
## 165  185.000     Chicago     West         4  5.220356    520  21      11   120
## 166  245.000 Minneapolis     West         6  5.501258    193  24       5    47
## 168  235.000       Texas     West        17  5.459586    213   3       4    61
## 169  425.000     Oakland     West         5  6.052089    441  76      11   113
## 171  165.000  California     West         1  5.105945    593  57      15   172
## 173  385.000     Chicago     West         6  5.953243    475  52       7   126
## 175 1000.000 Kansas City     West        11  6.907755    631  31       3   170
##     nHome nRuns
## 1       1    30
## 5      17    49
## 6      21   107
## 7       0    30
## 13     18    73
## 14      4    92
## 17     17    83
## 19      0    19
## 20      5    31
## 21      1    15
## 22      7    32
## 23      7    19
## 24      2    28
## 25      8    34
## 26     12    46
## 28     15    48
## 30     12    45
## 31      3    24
## 32     25    98
## 33     24    58
## 36     31    93
## 38     20    67
## 39      1    44
## 41     29    78
## 42     26    86
## 43      6    57
## 47     31   117
## 50     16    44
## 54      7    35
## 56     24    90
## 57     17    61
```

```
## 58       9    69
## 59       2    41
## 60      16    48
## 61      23    67
## 62       8    15
## 64      31   101
## 65       4    19
## 70       3    30
## 78      40   107
## 79       6    48
## 80       1    33
## 83      29   108
## 85       8    18
## 86      10    80
## 87       7    58
## 88      13    32
## 96      20    98
## 97      11    54
## 100     13    28
## 102     28    84
## 103     21    69
## 107      8    33
## 108     11    77
## 109     21    89
## 110     22    53
## 112     18    42
## 114     20    95
## 115      4    94
## 117     14    64
## 118     18    68
## 121     28    71
## 124      9    43
## 129      9    62
## 131      6    41
## 132      6    61
## 135      4    22
## 136     33    75
## 137     13    42
## 138     16    49
## 139      9    45
## 140     28   130
## 145      5    35
## 146      8    31
## 147     11    50
## 148      2    32
## 152      9    82
## 159     11    40
## 160     17    88
## 161      3    42
## 163     10    91
## 167      1    19
## 170      8   107
```

```
## 172      5     76
## 174      9     85
## 2       18     66
## 3        4     74
## 4        0     24
## 8        3     20
## 9        2     24
## 10       6     57
## 11       3     55
## 12       7     48
## 15      20     90
## 16       9     42
## 18       6     23
## 27      14     42
## 29      19     80
## 34       0     27
## 35       4     25
## 37       5     29
## 40      26     69
## 44      15     59
## 45       4     37
## 46      35     70
## 48       7     23
## 49       9     50
## 51      12     21
## 52      13     67
## 53      13     39
## 55      25     76
## 63      22     76
## 66      16     70
## 67      34     91
## 68      12     63
## 69      14     45
## 71       5     93
## 72      16     45
## 73      13     37
## 74       5     54
## 75      21     72
## 76       7     22
## 77       1     46
## 81       2     65
## 82      33     85
## 84       0     38
## 89       5     32
## 90       4     26
## 91       3     56
## 92       9     35
## 93      27     83
## 94       8     24
## 95       2     13
## 98       5     23
## 99      12     41
```

```
## 101      4    17
## 104     29    85
## 105     24    69
## 106     31   119
## 111     28    67
## 113      8    80
## 116     19    77
## 119      3    40
## 120      3    25
## 122      8    28
## 123     10    26
## 125      2    58
## 126     18   105
## 127     12    88
## 128     30    82
## 130     23    86
## 133      5    35
## 134      7    50
## 141     18    65
## 142     17    73
## 143     21    42
## 144      1    42
## 149     16    50
## 150     20    59
## 151      3    25
## 153     29    54
## 154      5    20
## 155     18    54
## 156      3    82
## 157      8    53
## 158      1    67
## 162     23    69
## 164      7    36
## 165     17    53
## 166     10    21
## 168      4    17
## 169      5    76
## 171     22    82
## 173      3    61
## 175      9    77
```

```
subdata <-
  data %>%
  group_by(Division) %>%
  summarize(c(total_count = n()), (mean_hits = mean(nHits)), (sd_hits = sd(nHits)), (se_
hits = sd(nHits)/sqrt(n())), (min_hits = min(nHits)), (max_hits = max(nHits)))
print(subdata)
```

```
## # A tibble: 2 × 7
##   Division `c(total_count = n())` (mean_hits = mean(nHi…¹ (sd_hits = sd(nHits)…²
##   <chr>                      <int>                      <dbl>                  <dbl>
## 1 East                          85                       113.                   48.8
## 2 West                          90                       102.                   41.2
## # ℹ abbreviated names: ¹`(mean_hits = mean(nHits))`, ²`(sd_hits = sd(nHits))`
## # ℹ 3 more variables: `(se_hits = sd(nHits)/sqrt(n()))` <dbl>,
## #   `(min_hits = min(nHits))` <int>, `(max_hits = max(nHits))` <int>
```

Assuming Normal Data

Pooled T-test

```
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns))
t.test_res <- t.test(data$nHits ~ data$Division, var.equal = TRUE)
print(t.test_res)
```

```
##
##  Two Sample t-test
##
## data:  data$nHits by data$Division
## t = 1.6116, df = 173, p-value = 0.1089
## alternative hypothesis: true difference in means between group East and group West is
not equal to 0
## 95 percent confidence interval:
##  -2.466278 24.413990
## sample estimates:
## mean in group East mean in group West
##           113.3294           102.3556
```

Satterthwaite T-test

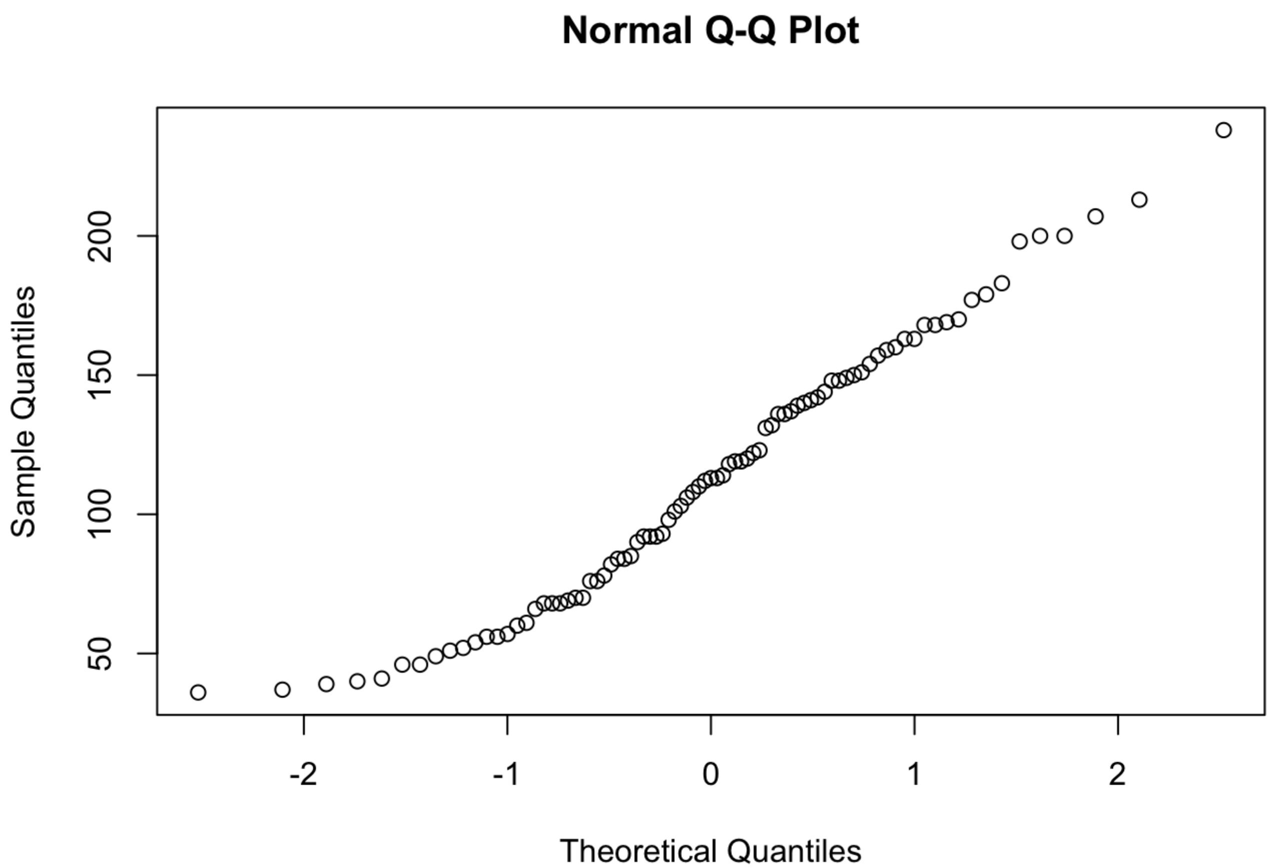
```
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns))
t.test(data$nHits ~ data$Division)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$nHits by data$Division
## t = 1.6038, df = 164.7, p-value = 0.1107
## alternative hypothesis: true difference in means between group East and group West is not equal to 0
## 95 percent confidence interval:
##  -2.536073 24.483786
## sample estimates:
## mean in group East mean in group West
##           113.3294           102.3556
```

Plots

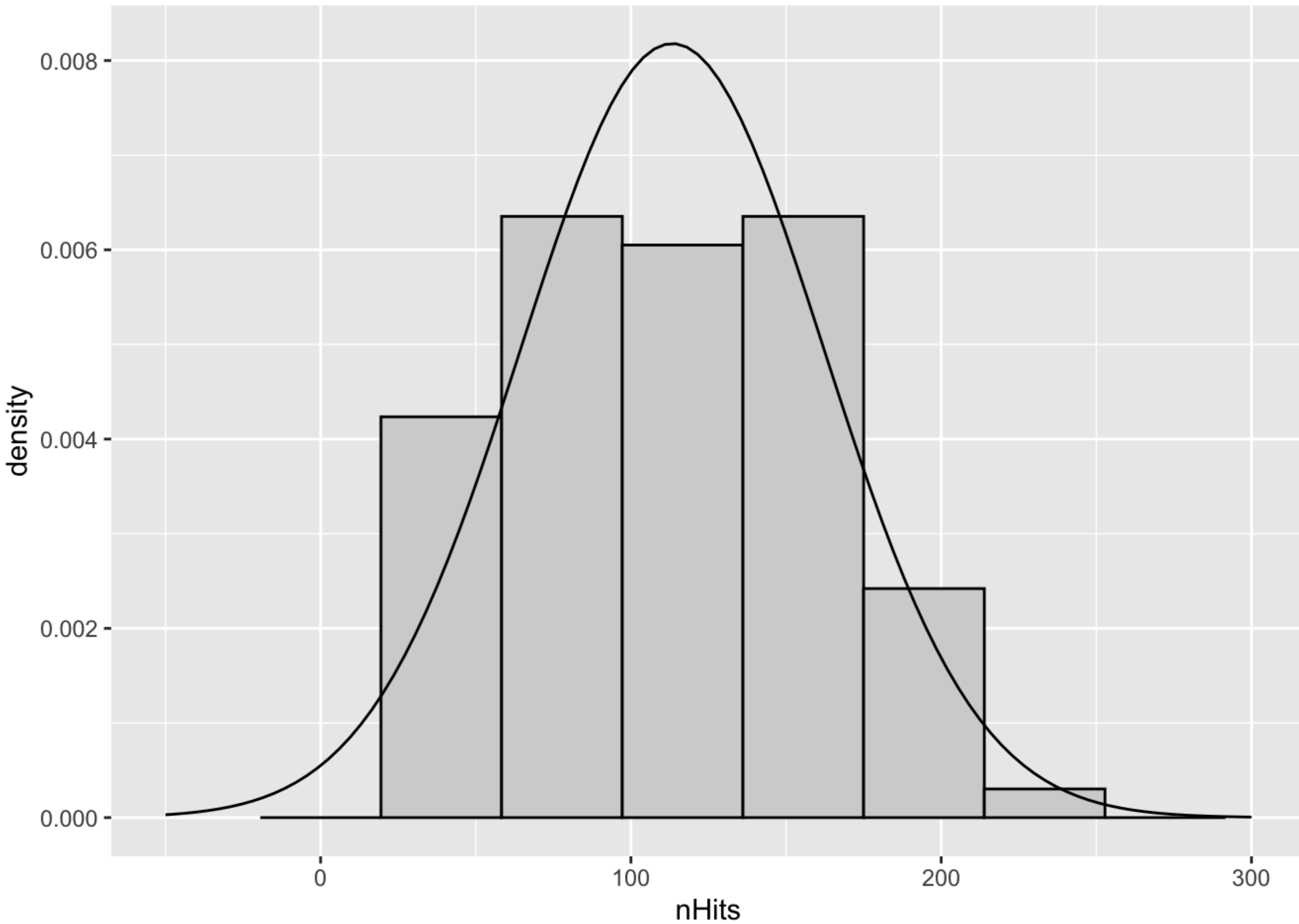
```
library("ggplot2")
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB, nError, nHits, nHome, nRuns))
eastsubdata <- filter(data, Division == 'East')
westsubdata <- filter(data, Division == 'West')

# East Plots
qqnorm(eastsubdata$nHits)
```

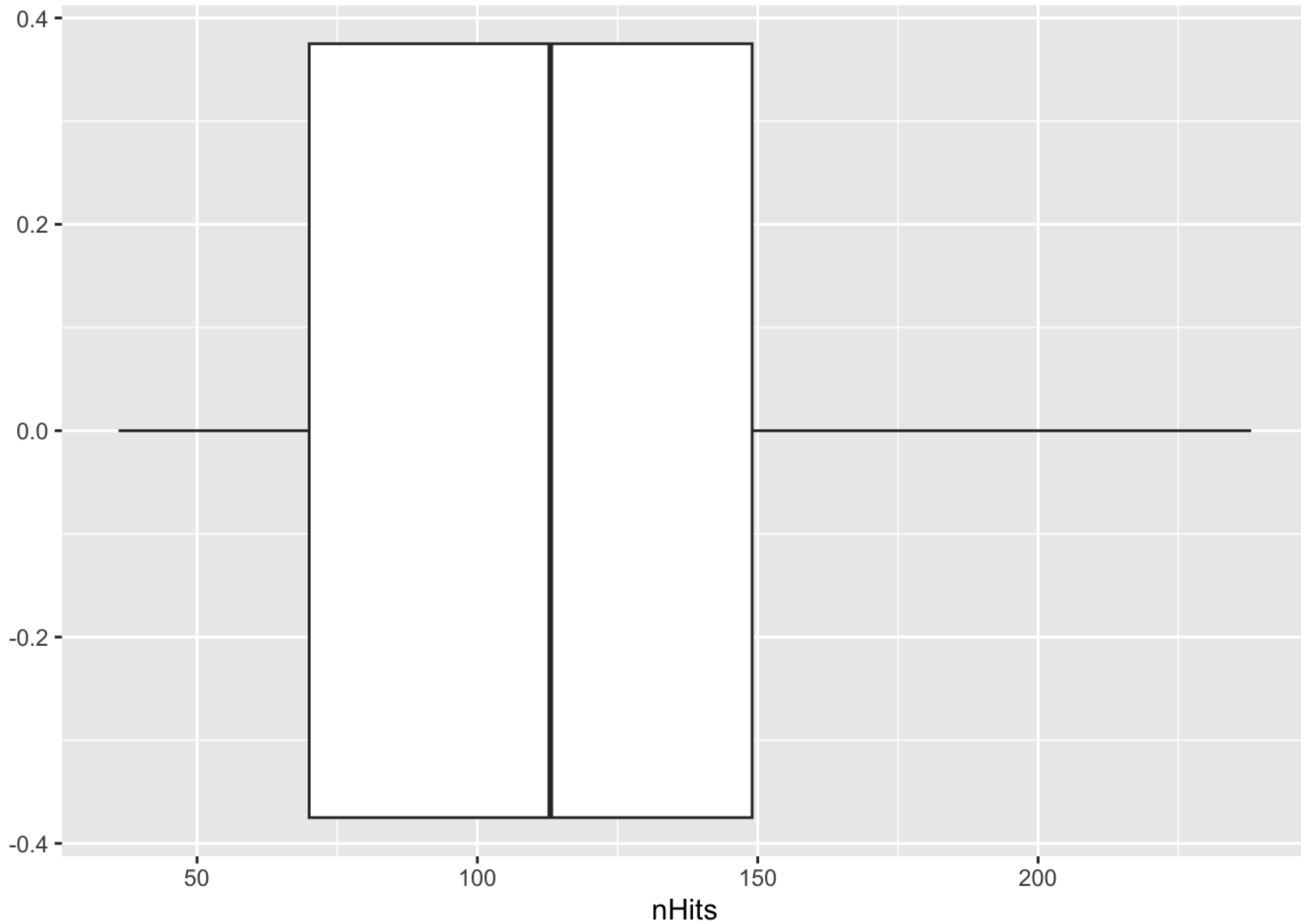



```
ggplot((eastsubdata), aes(x=nHits)) +
  geom_histogram(aes(y = after_stat(density)), fill='lightgray', col='black', bins = 10)
+
  scale_x_continuous(limits = c(-50,300)) +
  stat_function(fun = dnorm, args = list(mean=mean(eastsubdata$nHits), sd=sd(eastsubdata
$nHits)))
```

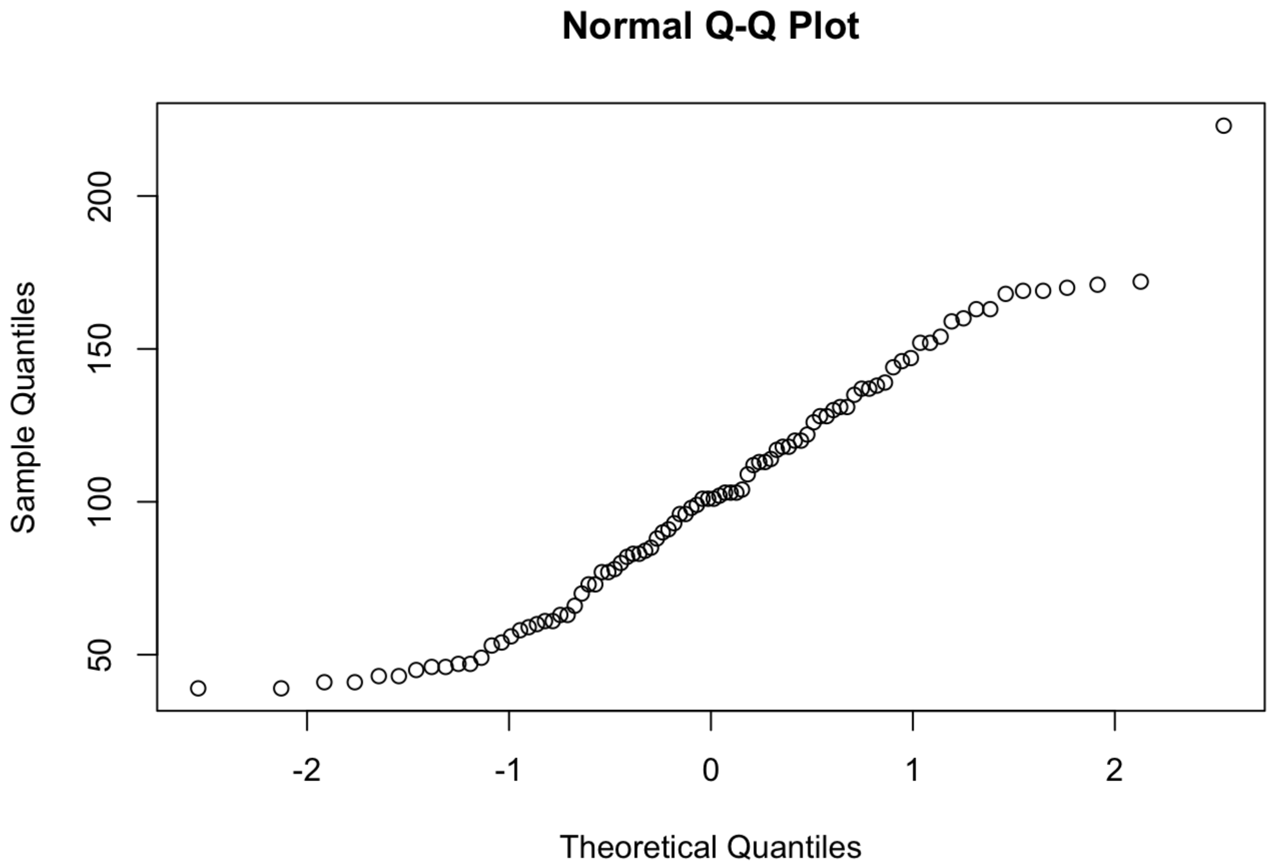
```
## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```



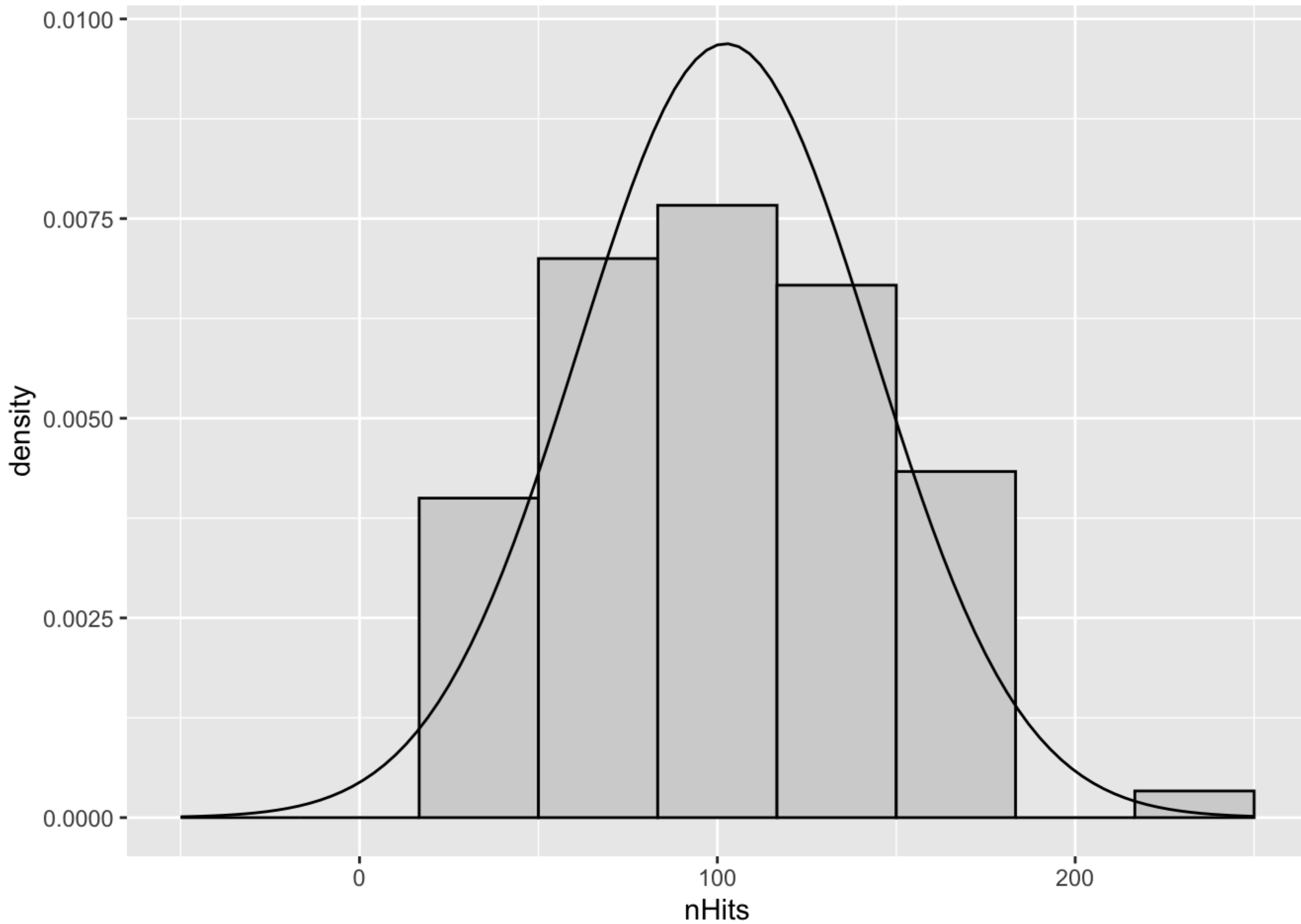

```
ggplot(eastsubdata, aes(x=nHits)) + geom_boxplot()
```



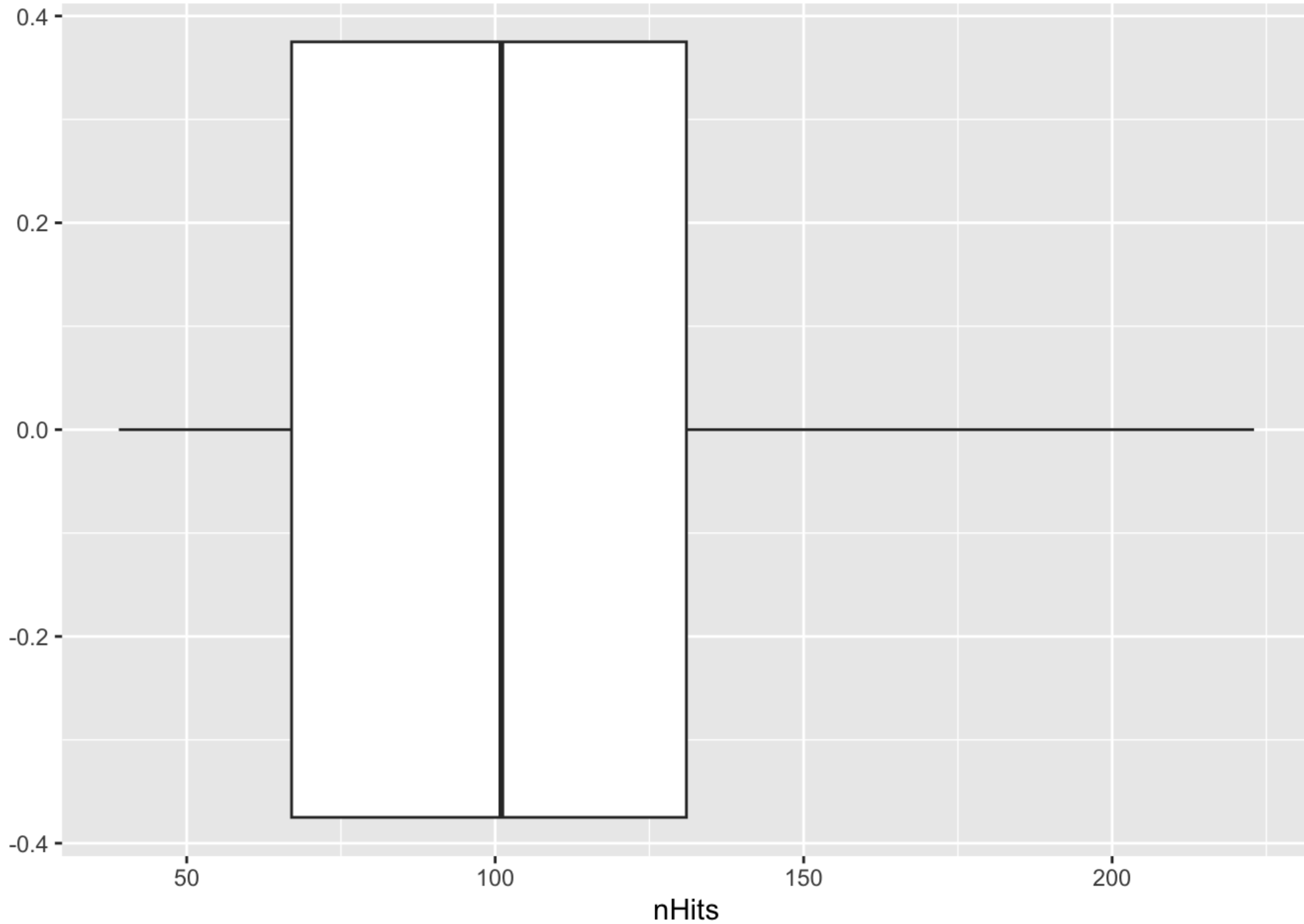

```
# West Plots
qqnorm(westsubdata$nHits)
```




```
ggplot((westsubdata), aes(x=nHits)) +
  geom_histogram(aes(y = after_stat(density)), fill='lightgray', col='black', bins = 10)
+
  scale_x_continuous(limits = c(-50,250)) +
  stat_function(fun = dnorm, args = list(mean=mean(westsubdata$nHits), sd=sd(westsubdata
$nHits)))
```




```
ggplot(westsubdata, aes(x=nHits)) + geom_boxplot()
```



Assuming Non-Normal Data

```
library("DescrTab2")
library("tidyverse")
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✔ tibble  3.2.1     ✔ purrr   0.3.4
## ✔ tidyr   1.2.0     ✔ stringr 1.4.1
## ✔ readr   2.1.2     ✔ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library("rstatix")
```

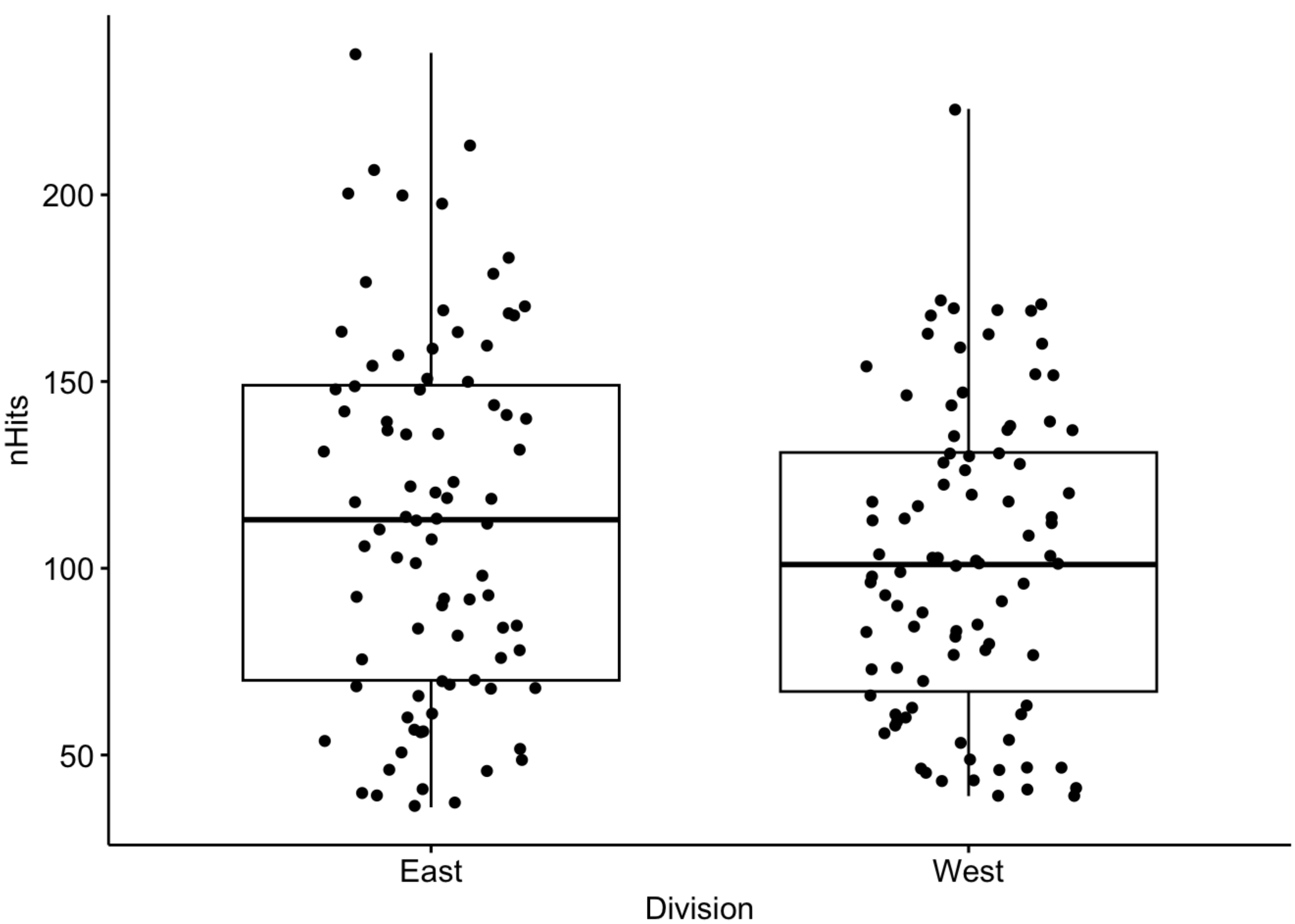
```
##
## Attaching package: 'rstatix'
##
## The following object is masked from 'package:stats':
##
##     filter
```

```
library("ggpubr")
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns))

data %>%
  group_by(Division) %>%
  get_summary_stats(nHits, type = "median_iqr")
```

```
## # A tibble: 2 × 5
##   Division variable     n median   iqr
##   <chr>    <chr>    <dbl>  <dbl> <dbl>
## 1 East     nHits       85    113    79
## 2 West     nHits       90    101    64
```

```
bxp <- ggboxplot(
  data, x = "Division", y = "nHits",
  ylab = "nHits", xlab = "Division", add = "jitter"
  )
bxp
```



```
stat.test <- data %>%
  wilcox_test(nHits ~ Division) %>%
  add_significance
stat.test
```

```
## # A tibble: 1 × 8
##   .y.   group1 group2    n1    n2 statistic     p p.signif
##   <chr> <chr>  <chr>  <int> <int>     <dbl> <dbl> <chr>
## 1 nHits East   West      85    90     4288. 0.167 ns
```

```
wilcox.test(data$nHits~data$Division)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data$nHits by data$Division
## W = 4288.5, p-value = 0.1669
## alternative hypothesis: true location shift is not equal to 0
```

```
kruskal.test(data$nHits~data$Division)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  data$nHits by data$Division
## Kruskal-Wallis chi-squared = 1.915, df = 1, p-value = 0.1664
```

```
ks.test(data$nHits~data$Division)
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  data$nHits by data$Division
## D = 0.14379, p-value = 0.2608
## alternative hypothesis: two-sided
```

Summary for Problem 1

```
# There was little reason to believe that there is a difference in the mean and median n
umber of hits for each division. Put differently, the p-value for the t-test assuming no
rmal data was not significant. This means that the assumption of normality is likely oka
y. Overall, there was not sufficient evidence to claim the mean and median differ betwee
n division.
```

Problem 2

```r
library("mosaic")
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2
```

```
##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected
by this.
```

```
##
## Attaching package: 'mosaic'
```

```
## The following object is masked from 'package:Matrix':
##
##     mean
```

```
## The following objects are masked from 'package:rstatix':
##
##     cor_test, prop_test, t_test
```

```
## The following object is masked from 'package:purrr':
##
##     cross
```

```
## The following object is masked from 'package:ggplot2':
##
##     stat
```

```
## The following objects are masked from 'package:arsenal':
##
##     iqr, relrisk
```

```
## The following objects are masked from 'package:dplyr':
##
##     count, do, tally
```

```
## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var
```

```
## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```
library("ggplot2")

datb <- filter(dat, League == 'National' | Division == 'East')
dim(datb)
```

```
## [1] 232  28
```

```
new_df = subset(datb, select = c(CrAtBat, CrBB, CrHits, CrHits2,CrRbi, CrRuns, Div, Sala
ry, YrMajor, nBB, nHits, nHome));

x = datb$nHits
y = (datb$CrHits)/(datb$YrMajor)

favstats(x)
```

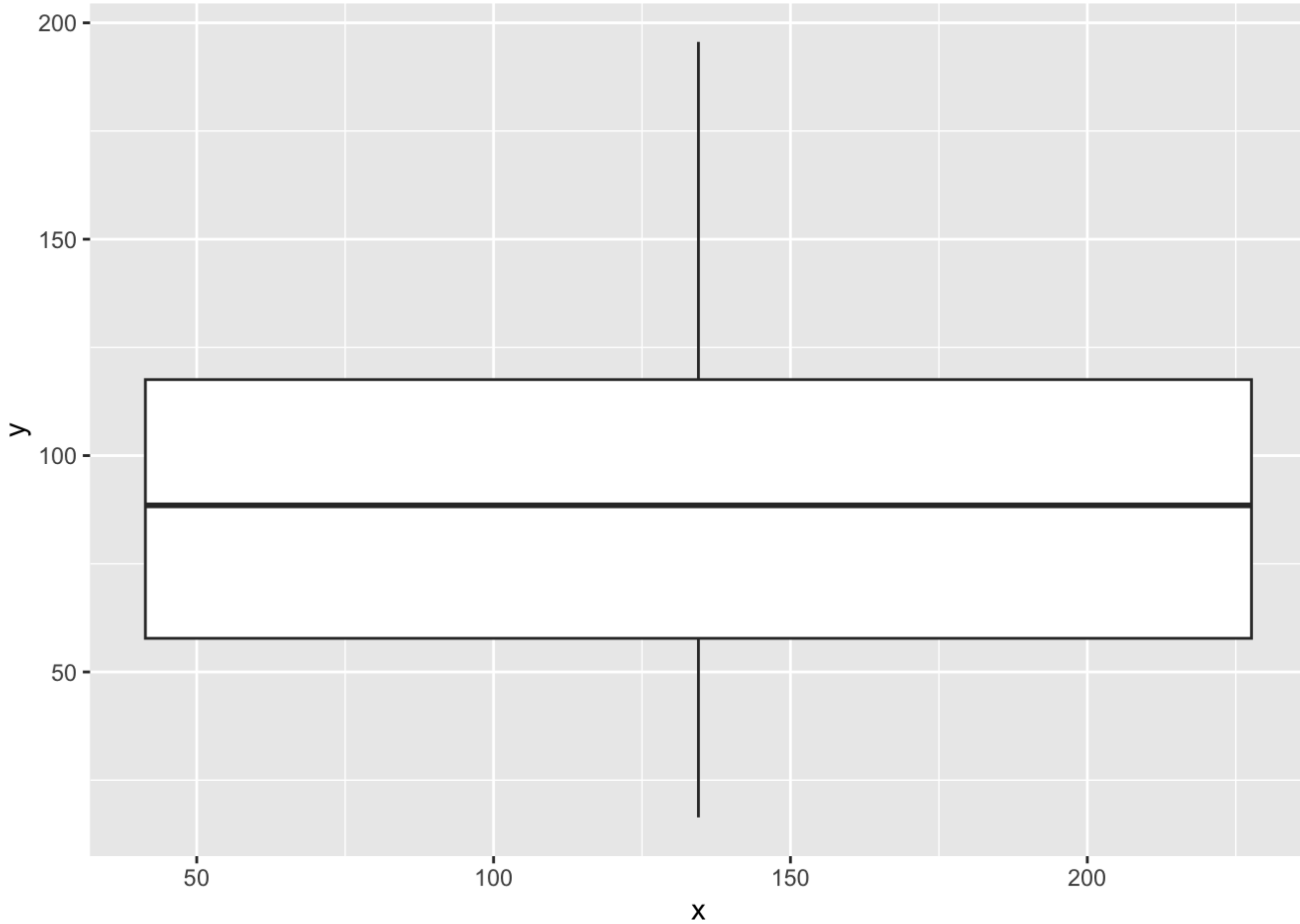
```
##  min Q1 median  Q3 max     mean       sd   n missing
##   31 68     96 140 238 103.8017 45.37286 232       0
```

```
favstats(y)
```

```
##   min     Q1 median       Q3   max     mean       sd   n missing
##  16.4 57.775   88.5 117.5625 195.6 89.18716 37.39127 232       0
```

```
ggplot(datb, aes(x=x, y=y)) + geom_boxplot()
```

```
## Warning: Continuous x aesthetic
## ℹ did you forget `aes(group = ...)`?
```




```
t.test(x,y)
```

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = 3.7861, df = 445.72, p-value = 0.000174
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   7.02840 22.20073
## sample estimates:
## mean of x mean of y
## 103.80172  89.18716
```

```
wilcox.test(x,y)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 31464, p-value = 0.001624
## alternative hypothesis: true location shift is not equal to 0
```

```
ks.test(x,y)
```

```
## Warning in ks.test.default(x, y): p-value will be approximate in the presence of
## ties
```

```
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  x and y
## D = 0.17672, p-value = 0.001426
## alternative hypothesis: two-sided
```

Summary for Problem 2

```
# There is a difference in the mean and median number of hits when comparing the 1986 na
tional league east team with the average career number. The p-value for the t-test is si
gnificant indicating that the difference in means is not equal to zero. Further, the p-v
alue of the wilcoxon test is significant indicating that the true location shift is not
equal to zero.
```

New Variable - nRuns

Problem 1

Subgroups

```
library("dplyr")
infield <- filter(dat, Position == '1B' | Position == '2B' | Position == 'SS' | Position
== '3B')
outfield <- filter(dat, Position == 'CF' | Position == 'RF' | Position == 'LF' | Positio
n == 'OF')
catcher <- filter(dat, Position == 'C')
CrOuts2 <- dat$CrOuts*dat$CrOuts
```

Plot League by Division

```
library("arsenal")
tab.lbyd <- table(dat$League, dat$Division)
lbyd <- freqlist(tab.lbyd)
summary(lbyd)
```

```
##
##
## |Var1     |Var2 | Freq| Cumulative Freq| Percent| Cumulative Percent|
## |:--------|:----|----:|---------------:|-------:|------------------:|
## |American |East |   85|              85|   26.40|              26.40|
## |         |West |   90|             175|   27.95|              54.35|
## |National |East |   72|             247|   22.36|              76.71|
## |         |West |   75|             322|   23.29|             100.00|
```

Plot Division for American Leagues

```
library("arsenal")
library("dplyr")
data <- filter(dat, League == 'American')
tab.adiv <- table(data$Division)
adiv <- freqlist(tab.adiv)
summary(adiv)
```

```
##
##
## |Var1 | Freq| Cumulative Freq| Percent| Cumulative Percent|
## |:----|----:|---------------:|-------:|------------------:|
## |East |   85|              85|   48.57|              48.57|
## |West |   90|             175|   51.43|             100.00|
```

Arrange by Division

```
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns, nRuns))
print(data[order(data$Division),])
```

```
##       Salary       Team Division YrMajor logSalary nAtBat nBB nError nHits
## 1         NA   Cleveland     East       1        NA    293  14     20    66
## 5   1100.000   Cleveland     East      13  7.003065    401  65      0    92
## 6    517.143     Detroit     East      10  6.248319    574  59     22   159
## 7    700.000   Baltimore     East       6  6.551080    239  22      6    60
## 13   776.667      Boston     East      18  6.655012    629  40     14   168
## 14   765.000   Cleveland     East       6  6.639876    587  70      3   163
## 17   612.500   Cleveland     East       5  6.417549    583  56     25   168
## 19        NA    New York     East       4        NA    161  17     12    36
## 20        NA   Milwaukee     East      16        NA    346  30      0    98
## 21    67.500   Milwaukee     East       2  4.212128    181  33      5    41
## 22   180.000   Milwaukee     East       4  5.192957    217   9      1    46
## 23        NA    New York     East      11        NA    194  30      2    40
## 24   305.000   Cleveland     East       6  5.720312    254  22      4    68
## 25   247.500   Cleveland     East       5  5.511411    205   9      4    57
## 26        NA   Milwaukee     East      16        NA    542  41      9   140
## 28        NA     Toronto     East      15        NA    336  52      0    84
## 30   675.000     Detroit     East      12  6.514713    403  39      5   101
## 31        NA   Milwaukee     East      14        NA    235  21      4    61
## 32  1350.000   Baltimore     East       6  7.207860    627  70     13   177
## 33    90.000   Cleveland     East       1  4.499810    416  16     10   113
## 36   950.000      Boston     East      17  6.856462    585  62      0   139
## 38   105.000     Detroit     East       4  4.653960    521  45     23   142
## 39        NA     Detroit     East      12        NA    419  44      1   113
## 41   535.000     Detroit     East      18  6.282267    507  91      2   122
## 42   933.333      Boston     East      15  6.838762    529  97      5   137
## 43   850.000     Toronto     East       9  6.745236    424  13      8   119
## 47  1975.000    New York     East       5  7.588324    677  53      6   238
## 50   110.000    New York     East       2  4.700480    280  47      2    82
## 54    70.000   Milwaukee     East       1  4.248495    317  32     26    78
## 56  1861.460    New York     East      14  7.529116    565  77      5   148
## 57  2460.000   Baltimore     East      10  7.807917    495  78     13   151
## 58        NA   Milwaukee     East       2        NA    524  54     20   132
## 59   375.000      Boston     East       8  5.926926    233  18     10    49
## 60        NA     Toronto     East      10        NA    395  35      7   106
## 61        NA   Baltimore     East      13        NA    397  53      4   114
## 62        NA   Baltimore     East       6        NA    210  15     15    37
## 64  1175.000     Toronto     East       5  7.069023    641  41     10   198
## 65    70.000   Milwaukee     East       1  4.248495    215  11     12    51
## 70   362.500     Toronto     East       8  5.893024    327  20     12    85
## 78  1237.500     Toronto     East       6  7.120848    589  69      3   170
## 79   430.000   Baltimore     East      15  6.063785    343  40     13   103
## 80        NA   Baltimore     East       5        NA    284  25      5    69
## 83   250.000   Cleveland     East       4  5.521461    663  32      6   200
## 85   275.000   Baltimore     East      14  5.616771    160  22      0    39
## 86   775.000   Cleveland     East       5  6.652863    599  32     18   183
## 87   850.000   Milwaukee     East      11  6.745236    497  26     10   136
## 88   365.000     Detroit     East      15  5.899897    210  28      0    70
## 96  2412.500      Boston     East      13  7.788419    618  62      8   200
## 97   300.000   Baltimore     East       6  5.703782    404  18      5    92
## 100       NA   Baltimore     East       2        NA    212  18      5    54
## 102 1300.000     Detroit     East       8  7.170120    441  68      2   118
```

```
## 103 1000.000    New York    East       14  6.907755    490  35        3   150
## 107  225.000     Detroit    East       12  5.416100    283  27        2    70
## 108  525.000   Baltimore    East       15  6.263398    491  37        2   141
## 109  787.500     Toronto    East        7  6.668863    589  64        6   149
## 110  800.000     Detroit    East       10  6.684612    327  38        6    84
## 112  145.000   Baltimore    East        3  4.976734    338  21        0    92
## 114  420.000     Detroit    East       10  6.040255    584  63       11   157
## 115  575.000      Boston    East        5  6.354370    625  65       14   179
## 117  700.000    New York    East       13  6.551080    490  49        0   148
## 118  550.000   Cleveland    East        6  6.309918    442  33        7   131
## 121  175.000    New York    East        3  5.164786    504  54       19   120
## 124  350.000   Baltimore    East        5  5.857933    369  49        6    93
## 129 1260.000   Milwaukee    East        9  7.138867    437  40       15   123
## 131  190.000     Detroit    East        5  5.247024    236  21        4    56
## 132  580.000   Cleveland    East        6  6.363028    473  29        9   154
## 135  250.000   Milwaukee    East       12  5.521461    216  15        4    56
## 136  215.000   Milwaukee    East        3  5.370638    466  72        8   108
## 137  400.000   Baltimore    East       18  5.991465    327  45        7    68
## 138       NA      Boston    East        7        NA    462  37        6   119
## 139  560.000    New York    East        9  6.327937    341  46        4   110
## 140 1670.000    New York    East        8  7.420579    608  89        6   160
## 145  250.000     Toronto    East        6  5.521461    246  13        1    76
## 146  400.000   Milwaukee    East       12  5.991465    205  17        2    52
## 147  450.000     Toronto    East       10  6.109248    348  43        6    90
## 148   70.000      Boston    East        1  4.248495    312  24       15    68
## 152 1000.000   Milwaukee    East       13  6.907755    522  62        1   163
## 159       NA      Boston    East       11        NA    425  24        8   112
## 160  530.000   Cleveland    East        8  6.272877    562  53       17   169
## 161  341.667     Detroit    East        8  5.833837    281  20        7    76
## 163  350.000     Toronto    East        4  5.857933    687  27       13   213
## 167       NA   Baltimore    East        5        NA    181  17        9    46
## 170 1600.000      Boston    East        5  7.377759    580 105       19   207
## 172  875.000    New York    East       12  6.774224    492  94       20   136
## 174  960.000     Toronto    East        8  6.866933    573  78       12   144
## 2    480.000     Seattle    West        3  6.173786    479  76       14   130
## 3    750.000     Oakland    West       11  6.620073    594  35       25   169
## 4    100.000 Kansas City    West        3  4.605170    298   7        9    73
## 8         NA Minneapolis    West        3        NA    183  11        0    39
## 9    175.000 Kansas City    West        5  5.164786    190  15       16    46
## 10        NA     Oakland    West       12        NA    407  65        9   104
## 11   115.000     Chicago    West        1  4.744932    426  62        2   109
## 12        NA  California    West       15        NA    442  43       11    98
## 15   900.000  California    West       14  6.802395    513  90        3   137
## 16        NA  California    West       17        NA    313  39        7    84
## 18   300.000     Seattle    West        7  5.703782    204  12        5    49
## 27   875.000     Chicago    West       17  6.774224    457  22        4   101
## 29  1200.000     Oakland    West        9  7.090077    591  39        4   168
## 34   230.000       Texas    West        4  5.438079    236  11       13    56
## 35        NA     Oakland    West       19        NA    242  27        0    58
## 37    75.000     Chicago    West        3  4.317488    199  21        5    53
## 40   850.000  California    West       14  6.745236    512  52       12   131
## 44   325.000     Seattle    West        6  5.783825    388  39        4   103
```

```
## 45   275.000     Oakland     West         4  5.616771     339   23         9    96
## 46        NA     Oakland     West        16        NA     561   33         8   118
## 48        NA Kansas City     West         5        NA     227   12         2    46
## 49   600.000     Oakland     West         9  6.396930     329   56         2    83
## 51   260.000       Texas     West        16  5.560682     155   22         1    41
## 52   475.000  California     West         4  6.163315     458   48        18   114
## 53   431.500       Texas     West         5  6.067268     314   16         4    83
## 55   145.000     Seattle     West         3  4.976734     511   61         8   138
## 63   750.000 Kansas City     West        14  6.620073     566   43        10   154
## 66  1500.000 Kansas City     West        14  7.313220     441   80        16   128
## 67   900.000 Minneapolis     West         6  6.802395     596   52        21   171
## 68   155.000 Minneapolis     West         4  5.043425     472   30        26   118
## 69   700.000  California     West        16  6.551080     283   26         5    77
## 71   400.000  California     West         5  5.991465     539   69         7   139
## 72        NA     Seattle     West        13        NA     315   58         0    59
## 73   500.000     Chicago     West         5  6.214608     282   29         5    78
## 74   600.000       Texas     West         8  6.396930     380   31         1   120
## 75   950.000     Chicago     West         7  6.856462     570   38         5   169
## 76   325.000 Kansas City     West        18  5.783825     278   18         0    70
## 77    87.500     Seattle     West         4  4.471639     445   29        16    99
## 81   100.000     Chicago     West         2  4.605170     438   71         9   103
## 82   165.000     Oakland     West         2  5.105945     600   65        14   144
## 84        NA     Chicago     West        10        NA     209   42         5    45
## 89        NA     Chicago     West        11        NA     225   26         0    61
## 90    95.000  California     West         2  4.553877     151   19         2    41
## 91    80.000     Seattle     West         5  4.382027     399   34         3   102
## 92        NA Kansas City     West        15        NA     336   23         0    93
## 93   200.000     Seattle     West         3  5.298317     616   32        15   163
## 94        NA Kansas City     West        12        NA     219   17         4    47
## 95    75.000 Minneapolis     West         3  4.317488     165   16         2    39
## 98   110.000     Chicago     West         4  4.700480     315   16         3    73
## 99   825.000 Kansas City     West        13  6.715383     429   57         4    91
## 101       NA     Oakland     West         3        NA     161   22         2    43
## 104 1310.000 Minneapolis     West         6  7.177782     550   71        10   147
## 105  300.000     Seattle     West         7  5.703782     344   88         0    85
## 106  365.000 Minneapolis     West         3  5.899897     680   34         6   223
## 111  587.500       Texas     West        13  6.375876     464   52         0   128
## 113       NA Kansas City     West         9        NA     508   46         9   146
## 116  780.000     Oakland     West         7  6.659294     489   34         9   131
## 119       NA Minneapolis     West         8        NA     317   19         4    88
## 120   68.000 Kansas City     West         1  4.219508     209   12         3    54
## 122  137.000 Minneapolis     West         3  4.919981     258   18         8    60
## 123  120.000     Oakland     West         3  4.787492     211   39         8    43
## 125  175.000     Chicago     West         2  5.164786     547   12        22   137
## 126  200.000       Texas     West         2  5.298317     572   65         3   152
## 127  750.000     Seattle     West         4  6.620073     526   77         1   163
## 128  172.000       Texas     West         1  5.147494     540   55        14   135
## 130       NA       Texas     West         5        NA     551   87        11   160
## 133  450.000  California     West        12  6.109248     271   33         3    77
## 134  300.000 Minneapolis     West         5  5.703782     357   39         4    96
## 141  487.500  California     West        20  6.189290     419   92         0   101
## 142       NA  California     West        11        NA     393   64         4    90
```

```
## 143  425.000     Chicago     West         5  6.052089    376  35       0    82
## 144       NA Kansas City     West         7        NA    307  29       2    80
## 149   97.500       Texas     West         1  4.579852    382  22       6   101
## 150  740.000 Minneapolis     West        12  6.606650    459  68       0   113
## 151  341.667  California     West        10  5.833837    288  16       7    63
## 153  100.000 Kansas City     West         6  4.605170    512  43      18   117
## 154   90.000     Seattle     West         3  4.499810    220  13       3    66
## 155  135.000       Texas     West         2  4.905275    461  35      11   112
## 156  475.000       Texas     West         6  6.163315    530  47      15   159
## 157  105.000 Minneapolis     West         2  4.653960    453  52       6   103
## 158  350.000     Seattle     West         4  5.857933    528  51      17   122
## 162  940.000 Minneapolis     West         6  6.845880    593  53       6   152
## 164       NA       Texas     West        17        NA    289  44       7    63
## 165  185.000     Chicago     West         4  5.220356    520  21      11   120
## 166  245.000 Minneapolis     West         6  5.501258    193  24       5    47
## 168  235.000       Texas     West        17  5.459586    213   3       4    61
## 169  425.000     Oakland     West         5  6.052089    441  76      11   113
## 171  165.000  California     West         1  5.105945    593  57      15   172
## 173  385.000     Chicago     West         6  5.953243    475  52       7   126
## 175 1000.000 Kansas City     West        11  6.907755    631  31       3   170
##     nHome nRuns nRuns.1
## 1       1    30      30
## 5      17    49      49
## 6      21   107     107
## 7       0    30      30
## 13     18    73      73
## 14      4    92      92
## 17     17    83      83
## 19      0    19      19
## 20      5    31      31
## 21      1    15      15
## 22      7    32      32
## 23      7    19      19
## 24      2    28      28
## 25      8    34      34
## 26     12    46      46
## 28     15    48      48
## 30     12    45      45
## 31      3    24      24
## 32     25    98      98
## 33     24    58      58
## 36     31    93      93
## 38     20    67      67
## 39      1    44      44
## 41     29    78      78
## 42     26    86      86
## 43      6    57      57
## 47     31   117     117
## 50     16    44      44
## 54      7    35      35
## 56     24    90      90
## 57     17    61      61
```

```
## 58       9    69      69
## 59       2    41      41
## 60      16    48      48
## 61      23    67      67
## 62       8    15      15
## 64      31   101     101
## 65       4    19      19
## 70       3    30      30
## 78      40   107     107
## 79       6    48      48
## 80       1    33      33
## 83      29   108     108
## 85       8    18      18
## 86      10    80      80
## 87       7    58      58
## 88      13    32      32
## 96      20    98      98
## 97      11    54      54
## 100     13    28      28
## 102     28    84      84
## 103     21    69      69
## 107      8    33      33
## 108     11    77      77
## 109     21    89      89
## 110     22    53      53
## 112     18    42      42
## 114     20    95      95
## 115      4    94      94
## 117     14    64      64
## 118     18    68      68
## 121     28    71      71
## 124      9    43      43
## 129      9    62      62
## 131      6    41      41
## 132      6    61      61
## 135      4    22      22
## 136     33    75      75
## 137     13    42      42
## 138     16    49      49
## 139      9    45      45
## 140     28   130     130
## 145      5    35      35
## 146      8    31      31
## 147     11    50      50
## 148      2    32      32
## 152      9    82      82
## 159     11    40      40
## 160     17    88      88
## 161      3    42      42
## 163     10    91      91
## 167      1    19      19
## 170      8   107     107
```

```
## 172      5     76        76
## 174      9     85        85
## 2       18     66        66
## 3        4     74        74
## 4        0     24        24
## 8        3     20        20
## 9        2     24        24
## 10       6     57        57
## 11       3     55        55
## 12       7     48        48
## 15      20     90        90
## 16       9     42        42
## 18       6     23        23
## 27      14     42        42
## 29      19     80        80
## 34       0     27        27
## 35       4     25        25
## 37       5     29        29
## 40      26     69        69
## 44      15     59        59
## 45       4     37        37
## 46      35     70        70
## 48       7     23        23
## 49       9     50        50
## 51      12     21        21
## 52      13     67        67
## 53      13     39        39
## 55      25     76        76
## 63      22     76        76
## 66      16     70        70
## 67      34     91        91
## 68      12     63        63
## 69      14     45        45
## 71       5     93        93
## 72      16     45        45
## 73      13     37        37
## 74       5     54        54
## 75      21     72        72
## 76       7     22        22
## 77       1     46        46
## 81       2     65        65
## 82      33     85        85
## 84       0     38        38
## 89       5     32        32
## 90       4     26        26
## 91       3     56        56
## 92       9     35        35
## 93      27     83        83
## 94       8     24        24
## 95       2     13        13
## 98       5     23        23
## 99      12     41        41
```

```
## 101     4    17      17
## 104    29    85      85
## 105    24    69      69
## 106    31   119     119
## 111    28    67      67
## 113     8    80      80
## 116    19    77      77
## 119     3    40      40
## 120     3    25      25
## 122     8    28      28
## 123    10    26      26
## 125     2    58      58
## 126    18   105     105
## 127    12    88      88
## 128    30    82      82
## 130    23    86      86
## 133     5    35      35
## 134     7    50      50
## 141    18    65      65
## 142    17    73      73
## 143    21    42      42
## 144     1    42      42
## 149    16    50      50
## 150    20    59      59
## 151     3    25      25
## 153    29    54      54
## 154     5    20      20
## 155    18    54      54
## 156     3    82      82
## 157     8    53      53
## 158     1    67      67
## 162    23    69      69
## 164     7    36      36
## 165    17    53      53
## 166    10    21      21
## 168     4    17      17
## 169     5    76      76
## 171    22    82      82
## 173     3    61      61
## 175     9    77      77
```

```
subdata <-
  data %>%
  group_by(Division) %>%
  summarize(c(total_count = n()), (mean_outs = mean(nRuns)), (sd_outs = sd(nRuns)), (se_
outs = sd(nRuns)/sqrt(n())), (min_outs = min(nRuns)), (max_outs = max(nRuns)))
print(subdata)
```

```
## # A tibble: 2 × 7
##   Division `c(total_count = n())` (mean_outs = mean(nRu…¹ (sd_outs = sd(nRuns)…²
##   <chr>                     <int>                   <dbl>                  <dbl>
## 1 East                         85                    58.5                   28.1
## 2 West                         90                    53.2                   23.9
## # ℹ abbreviated names: ¹`(mean_outs = mean(nRuns))`, ²`(sd_outs = sd(nRuns))`
## # ℹ 3 more variables: `(se_outs = sd(nRuns)/sqrt(n()))` <dbl>,
## #   `(min_outs = min(nRuns))` <int>, `(max_outs = max(nRuns))` <int>
```

Assuming Normal Data

Pooled T-test

```
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns, nRuns))
t.test_res <- t.test(data$nRuns ~ data$Division, var.equal = TRUE)
print(t.test_res)
```

```
##
##  Two Sample t-test
##
## data:  data$nRuns by data$Division
## t = 1.3531, df = 173, p-value = 0.1778
## alternative hypothesis: true difference in means between group East and group West is
not equal to 0
## 95 percent confidence interval:
##  -2.444213 13.101729
## sample estimates:
## mean in group East mean in group West
##           58.51765           53.18889
```

Satterthwaite T-test

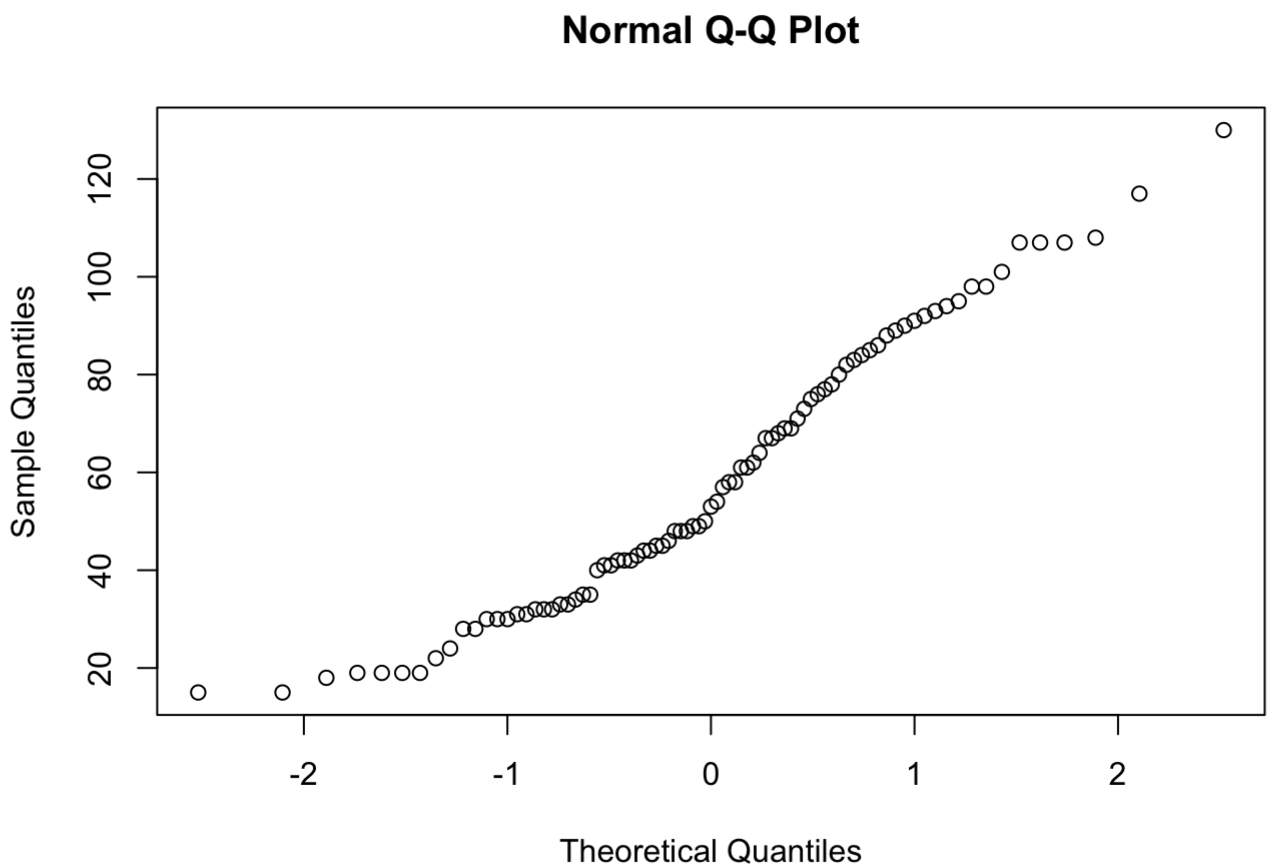
```
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns, nRuns))
t.test(data$nRuns ~ data$Division)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$nRuns by data$Division
## t = 1.3469, df = 165.29, p-value = 0.1798
## alternative hypothesis: true difference in means between group East and group West is
not equal to 0
## 95 percent confidence interval:
##  -2.482434 13.139951
## sample estimates:
## mean in group East mean in group West
##           58.51765           53.18889
```

Plots

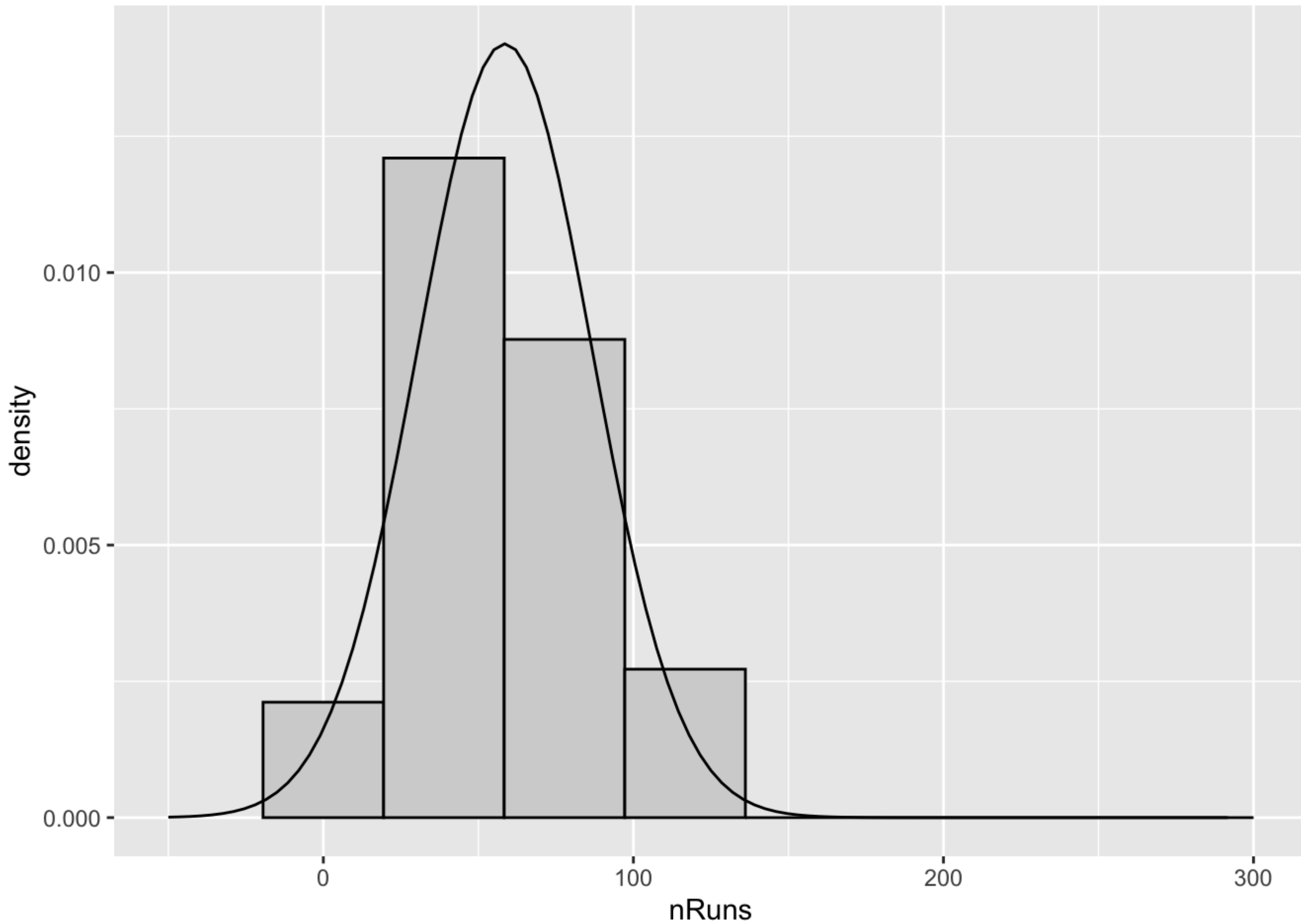
```
library("ggplot2")
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns, nRuns))
eastsubdata <- filter(data, Division == 'East')
westsubdata <- filter(data, Division == 'West')

# East Plots
qqnorm(eastsubdata$nRuns)
```

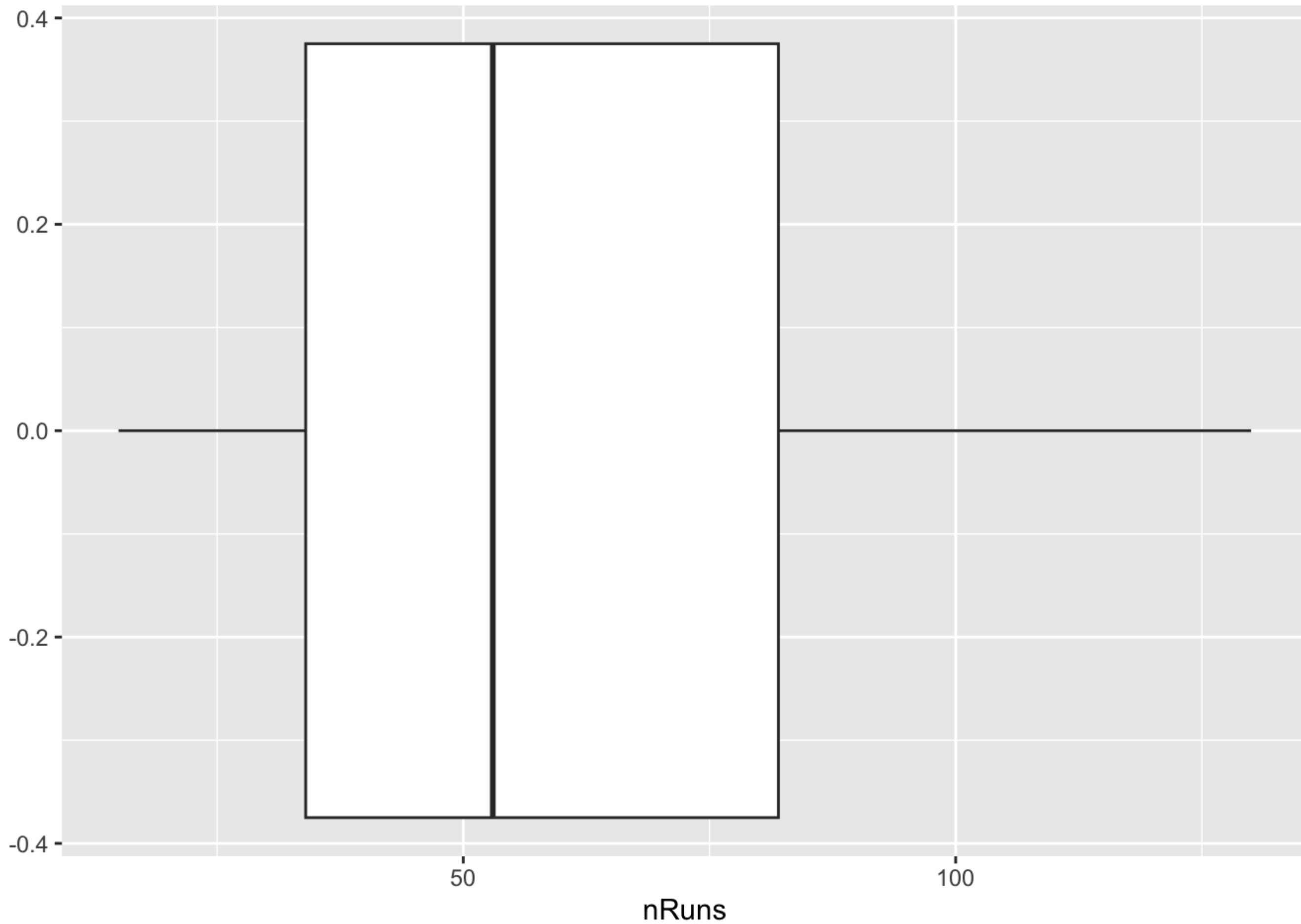



```
ggplot((eastsubdata), aes(x=nRuns)) +
  geom_histogram(aes(y = after_stat(density)), fill='lightgray', col='black', bins = 10)
+
  scale_x_continuous(limits = c(-50,300)) +
  stat_function(fun = dnorm, args = list(mean=mean(eastsubdata$nRuns), sd=sd(eastsubdata
$nRuns)))
```

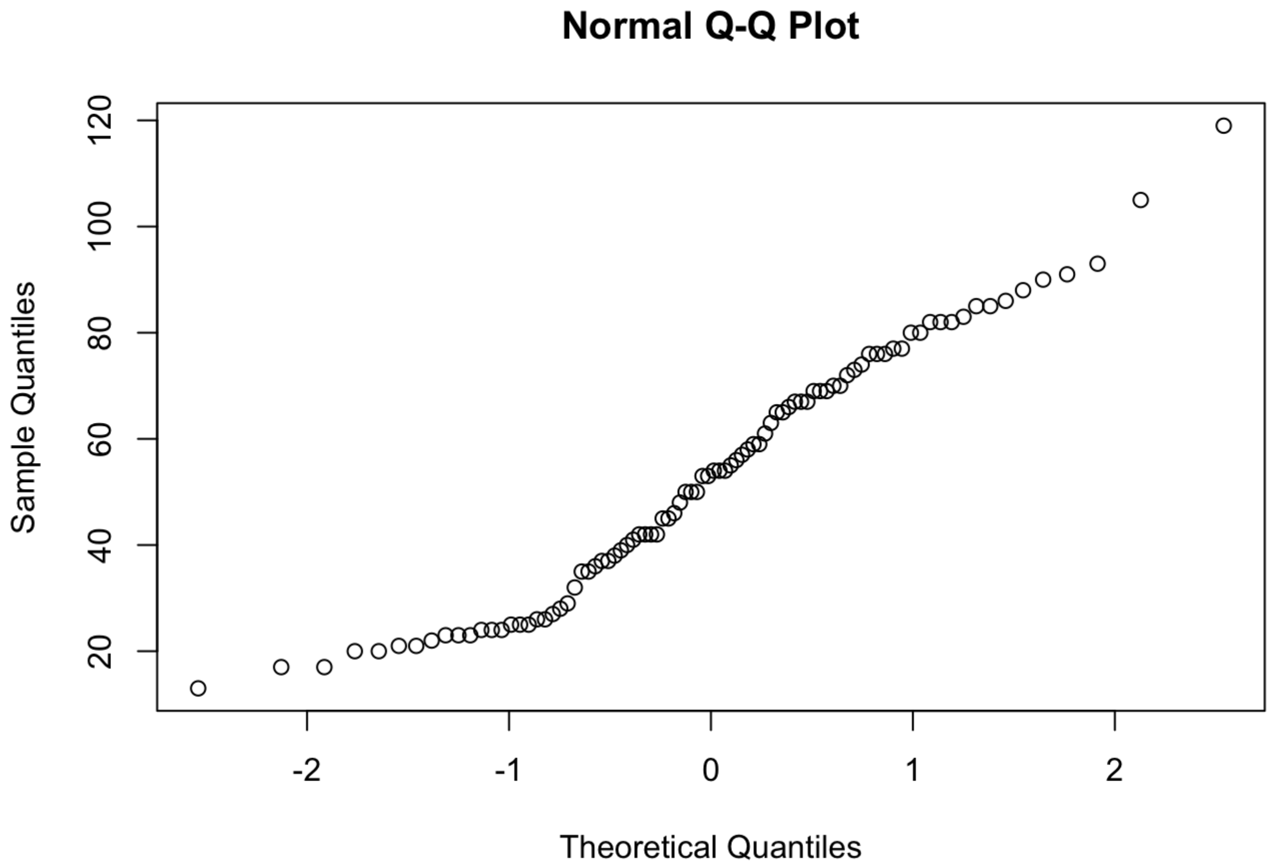
```
## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```



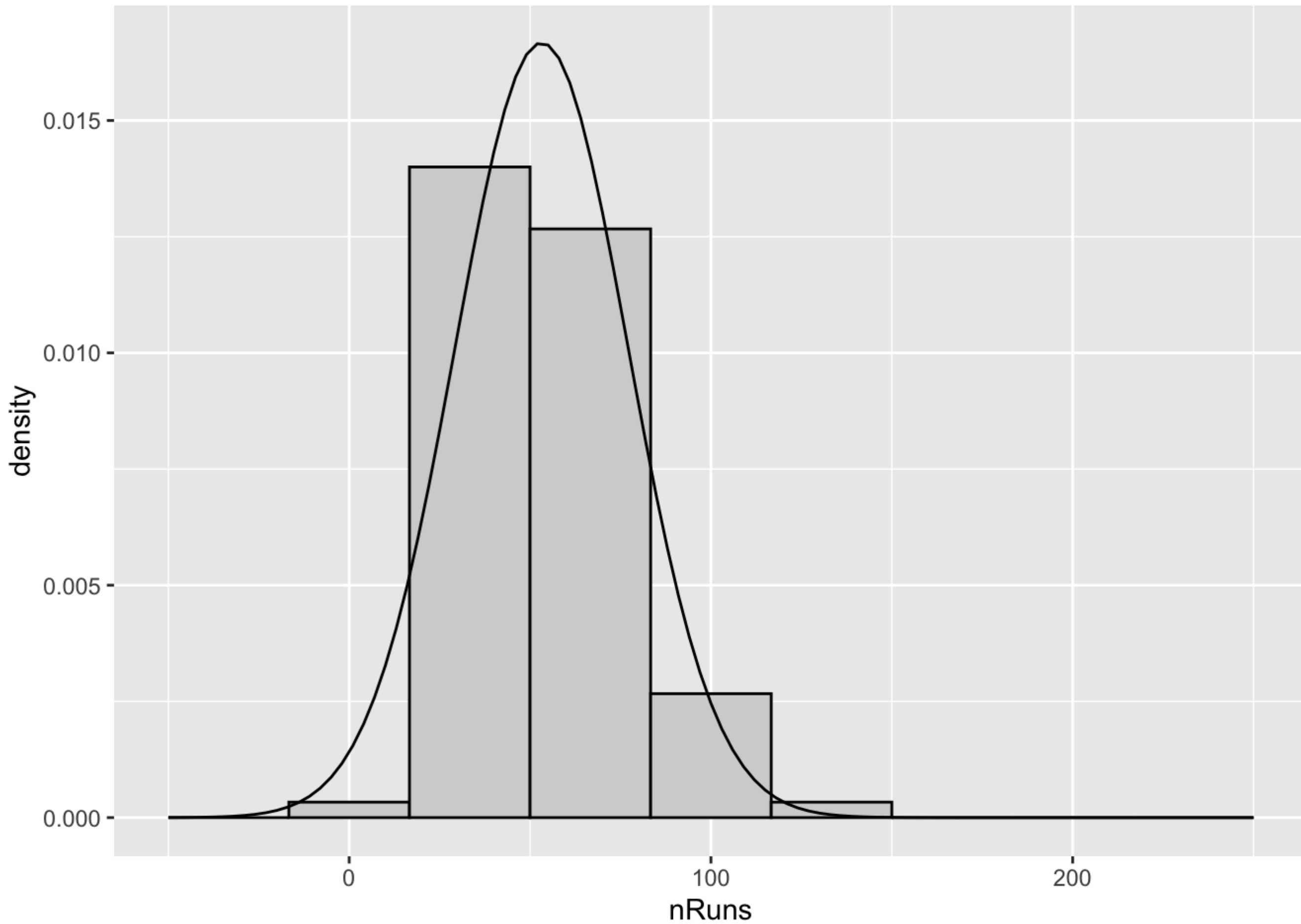

```
ggplot(eastsubdata, aes(x=nRuns)) + geom_boxplot()
```



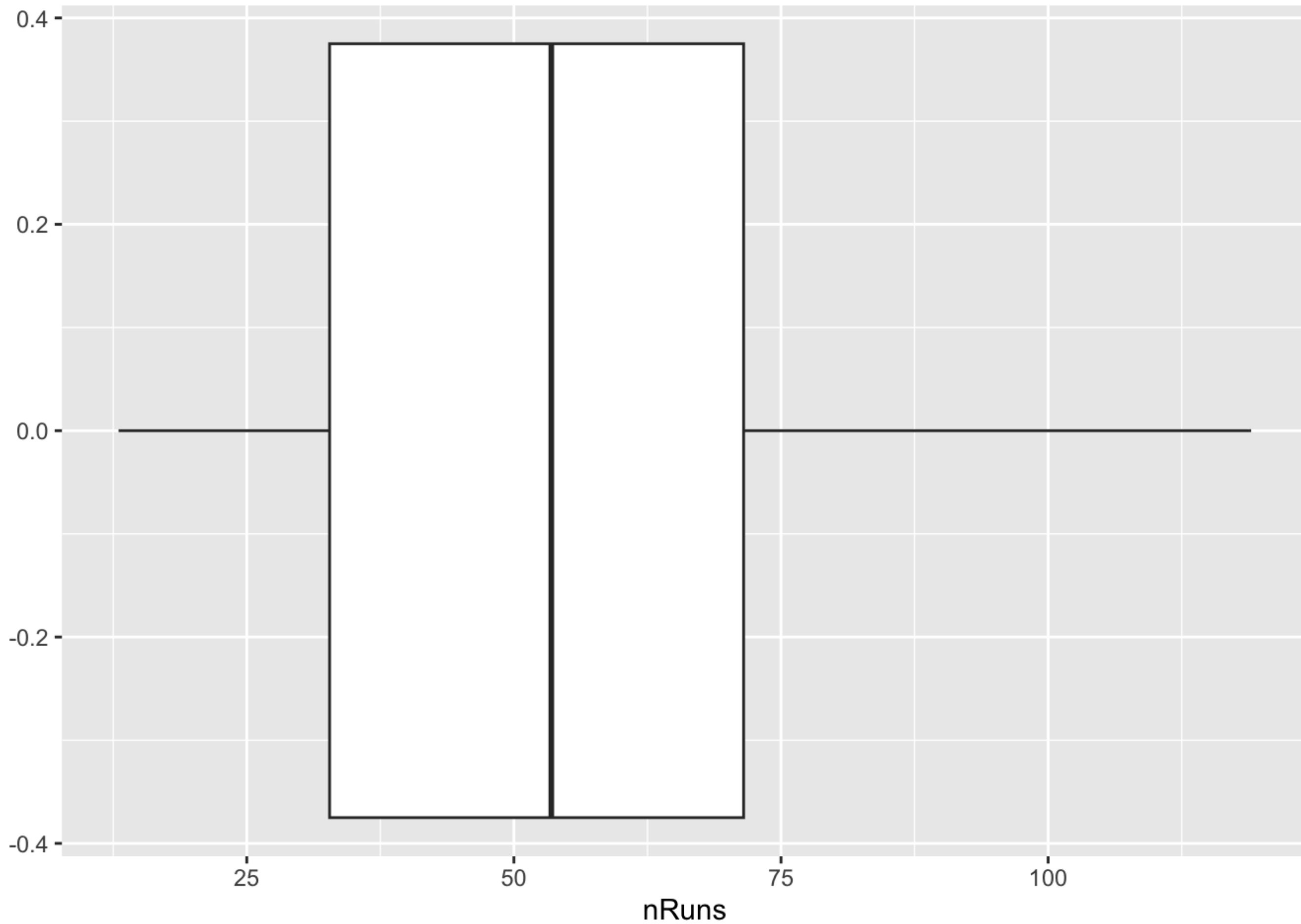

```
# West Plots
qqnorm(westsubdata$nRuns)
```




```
ggplot((westsubdata), aes(x=nRuns)) +
  geom_histogram(aes(y = after_stat(density)), fill='lightgray', col='black', bins = 10)
+
  scale_x_continuous(limits = c(-50,250)) +
  stat_function(fun = dnorm, args = list(mean=mean(westsubdata$nRuns), sd=sd(westsubdata
$nRuns)))
```




```
ggplot(westsubdata, aes(x=nRuns)) + geom_boxplot()
```



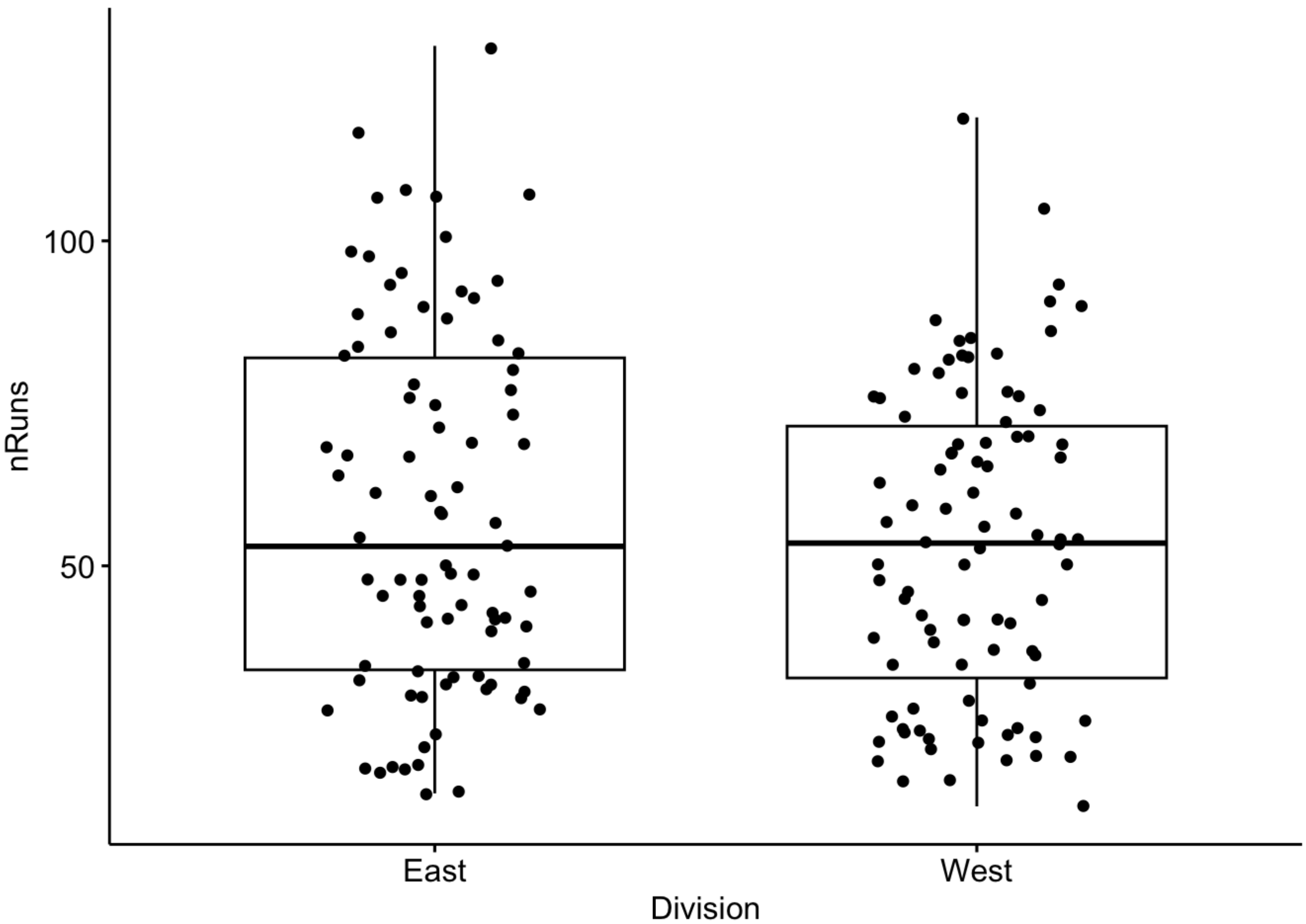
Assuming Non-Normal Data

```
library("DescrTab2")
library("tidyverse")
library("rstatix")
library("ggpubr")
data <- filter(dat, League == 'American')
data <- subset(data, select = c(Salary, Team, Division, YrMajor, logSalary, nAtBat, nBB,
nError, nHits, nHome, nRuns, nRuns))

data %>%
  group_by(Division) %>%
  get_summary_stats(nRuns, type = "median_iqr")
```

```
## # A tibble: 2 × 5
##   Division variable     n median   iqr
##   <chr>    <chr>    <dbl>  <dbl> <dbl>
## 1 East     nRuns       85   53    48
## 2 West     nRuns       90   53.5  38.8
```

```
bxp <- ggboxplot(
  data, x = "Division", y = "nRuns",
  ylab = "nRuns", xlab = "Division", add = "jitter"
  )
bxp
```




```
stat.test <- data %>%
  wilcox_test(nRuns ~ Division) %>%
  add_significance
stat.test
```

```
## # A tibble: 1 × 8
##   .y.   group1 group2    n1    n2 statistic     p p.signif
##   <chr> <chr>  <chr>  <int> <int>     <dbl> <dbl> <chr>
## 1 nRuns East   West      85    90      4201 0.262 ns
```

```
wilcox.test(data$nRuns~data$Division)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data$nRuns by data$Division
## W = 4201, p-value = 0.2622
## alternative hypothesis: true location shift is not equal to 0
```

```
kruskal.test(data$nRuns~data$Division)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  data$nRuns by data$Division
## Kruskal-Wallis chi-squared = 1.2604, df = 1, p-value = 0.2616
```

```
ks.test(data$nRuns~data$Division)
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  data$nRuns by data$Division
## D = 0.13595, p-value = 0.3119
## alternative hypothesis: two-sided
```

Summary for Problem 1

```
# There was little reason to believe that there is a difference in the mean and median n
umber of runs for each division. Put differently, the p-value for the t-test assuming no
rmal data was not significant. This means that the assumption of normality is likely oka
y. Overall, there was not sufficient evidence to claim the mean and median differ betwee
n division.
```

Problem 2

```
library("mosaic")
library("ggplot2")

datb <- filter(dat, League == 'National' | Division == 'East')
dim(datb)
```

```
## [1] 232  28
```

```
new_df = subset(datb, select = c(CrAtBat, CrBB, CrHits, CrHits2,CrRbi, CrRuns, Div, Sala
ry, YrMajor, nBB, nHits, nHome, nRuns));

x = datb$nRuns
y = (datb$CrRuns)/(datb$YrMajor)

favstats(x)
```

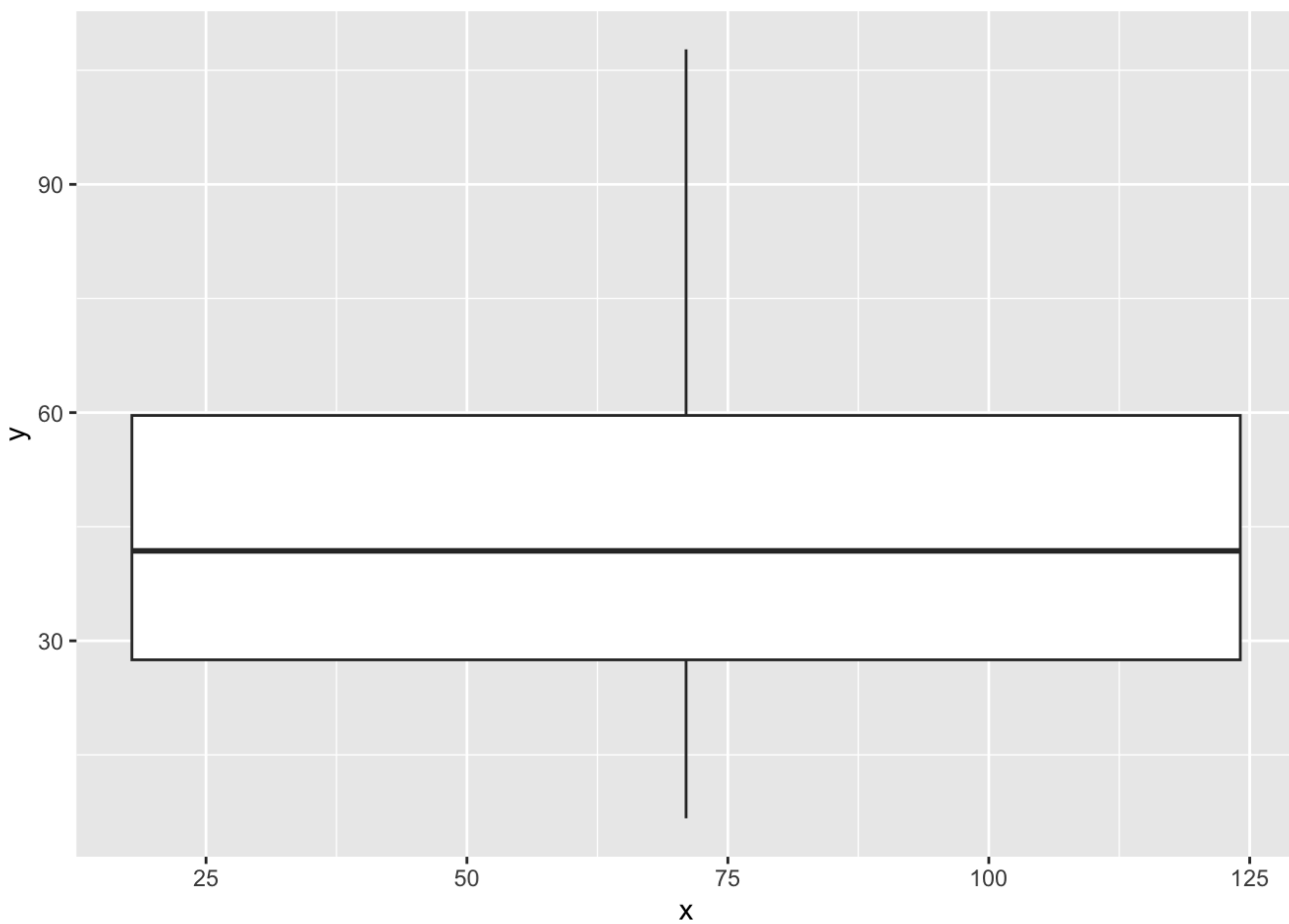
```
##  min    Q1 median Q3 max     mean       sd   n missing
##   12 31.75     48 69 130 51.84052 25.52184 232       0
```

```
favstats(y)
```

```
##       min   Q1   median       Q3    max     mean       sd   n missing
##  6.666667 27.5 41.81944 59.62692 107.75 44.66428 20.69896 232       0
```

```
ggplot(datb, aes(x=x, y=y)) + geom_boxplot()
```

```
## Warning: Continuous x aesthetic
## ℹ did you forget `aes(group = ...)`?
```



```
t.test(x,y)
```

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = 3.3263, df = 443.12, p-value = 0.0009532
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   2.936233 11.416232
## sample estimates:
## mean of x mean of y
##  51.84052  44.66428
```

```
wilcox.test(x,y)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 30642, p-value = 0.009798
## alternative hypothesis: true location shift is not equal to 0
```

```
ks.test(x,y)
```

```
## Warning in ks.test.default(x, y): p-value will be approximate in the presence of
## ties
```

```
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  x and y
## D = 0.14655, p-value = 0.01371
## alternative hypothesis: two-sided
```

Summary for Problem 2

```
# There is a difference in the mean and median number of runs when comparing the 1986 na
tional league east team with the average career number. The p-value for the t-test is si
gnificant indicating that the difference in means is not equal to zero. Further, the p-v
alue of the wilcoxon test is significant indicating that the true location shift is not
equal to zero.
```