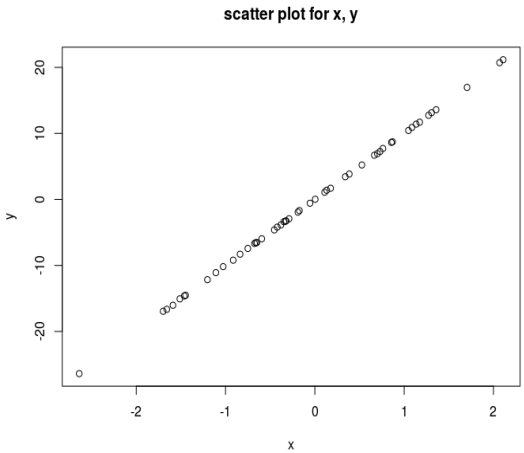
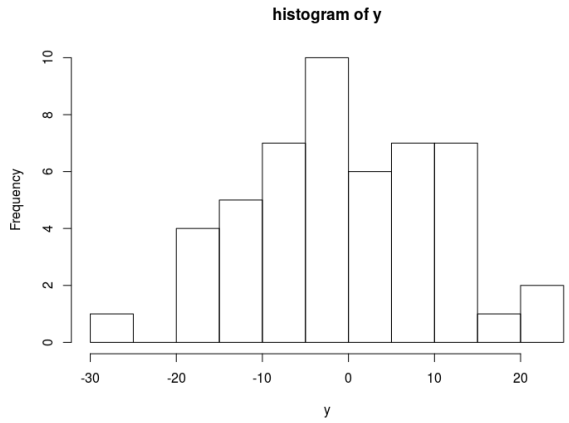
Violet Chang || chan1300 || 5197617

Stat 3022 HW1

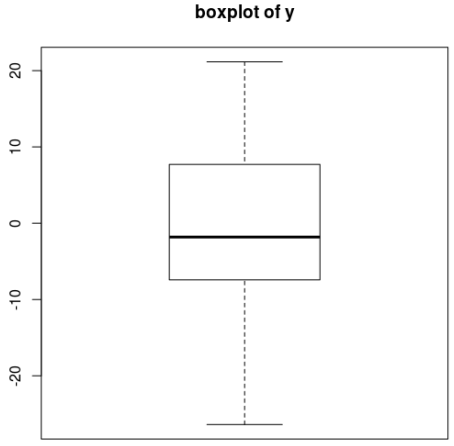
```
## Question 1
# Question 1: (a)
> x = rnorm(50)
> x
 [1] -1.588757874 -0.173113730  1.048441196 -2.638396690  0.176449139  2.108430792  1.357605774
-1.657939534 -0.750492198 -0.675130223 -0.454274591
[12] -0.648333712 -0.661346920 -1.446911931  2.068834049  0.004292554  0.109610191  0.340529607
1.306971919  0.666155648  1.273373534 -0.288888988
[23] -0.837211404 -0.594396485 -1.108104163 -0.327771359 -0.346929680 -0.420436175  0.857092254
0.699800475  0.526496189  0.869972042 -0.189712483
[34] -1.202659202 -1.462067757  1.134754340 -1.511790929  0.385456465  0.132545234 -0.054181594
1.704146151 -0.321814218 -0.914158701 -1.026541178
[45] -0.379825986 -1.698538632  1.087109573  0.729764847  1.174018079  0.762350212
> mean(x)
[1] -0.05711052
> sd(x)
[1] 1.067795
> median(x)
[1] -0.1814131
# Question 1: (b)
> e = rnorm(50, sd = 0.05)
> e
 [1] -0.148193689  0.038111008 -0.047617104 -0.001259398 -0.075550753  0.067966293  0.001059469
-0.066459491  0.079747868  0.082503245 -0.072191737
[12] -0.047727119  0.059258508 -0.050258971  0.015277697 -0.019408997 -0.015220947  0.027064631
0.065036213  0.027882416  0.007674216 -0.041812038
[23]  0.074910373 -0.021075638 -0.012408847 -0.008822217  0.093759345  0.008498428  0.061242041
-0.073037320 -0.045320520  0.016626681 -0.034637146
[34] -0.132891752 -0.002704490  0.053505821  0.056058806 -0.014682544  0.066437408 -0.035757009
-0.077900525 -0.037623419 -0.063056974  0.080865581
[45] -0.061978333  0.040481689  0.027223989 -0.040068363 -0.050138423  0.085646688
# Question 1: (c)
> y = 10 * x + e
> y
 [1] -16.03577243  -1.69302630  10.43679486 -26.38522630   1.68894064  21.15227421  13.57711721
-16.64585483  -7.42517411  -6.66879898  -4.61493765
[12]  -6.53106424  -6.55421069 -14.51937829  20.70361818   0.02351655   1.08088097   3.43236070
13.13475540   6.68943890  12.74140956  -2.93070192
[23]  -8.29720366  -5.96504049 -11.09345048  -3.28653580  -3.37553746  -4.19586332   8.63216458
6.92496743   5.21964137   8.71634710  -1.93176198
[34] -12.15948377 -14.62338206  11.40104922 -15.06185048   3.83988211   1.39188975  -0.57757294
16.96356099  -3.25576560  -9.20464398 -10.18454620
[45]  -3.86023819 -16.94490464  10.89831972   7.25758011  11.69004237   7.70914880
# Question 1: (d)
> plot(x, y, main  = "scatter plot for x, y", xlab = "x", ylab = "y")
```



```
# Question 1: (e)
> hist(y, main = "histogram of y")
```




```
> boxplot(y, main = "boxplot of y")
```



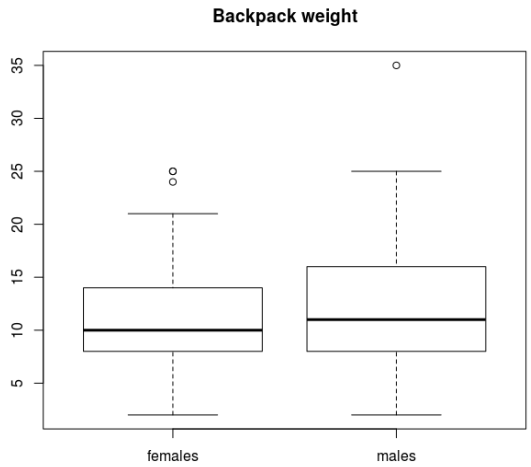

```
## Question 2
# Question 2: (a)
> library(Stat2Data)
> data(Backpack)
> backpack=data.frame(Backpack)
# Question 2: (b)
> max(backpack$BackpackWeight)
[1] 35
> min(backpack$BackpackWeight)
[1] 2
# Question 2: (c)
> Backpack_p = subset(backpack, Backpack$BackProblems == '1')
> View(Backpack_p)
> Backpack_p
   BackpackWeight BodyWeight     Ratio BackProblems     Major Year    Sex Status Units
1               9        125 0.0720000            1       Bio    3 Female      U    13
3              10        120 0.0833333            1       GRC    4 Female      U    14
8               4        185 0.0216216            1      ARCE    5 Female      U    18
14             11        170 0.0647059            1       CPE    4   Male      U    16
15             12        125 0.0960000            1        CS    3 Female      U    12
```

```
17             13        145 0.0896552            1       Bio    4 Female      U    16
19             10        125 0.0800000            1        LS    3 Female      U    17
22             14        117 0.1196580            1        EE    2 Female      U    12
26             18        160 0.1125000            1      Mate    1   Male      U    13
30              6        165 0.0363636            1 Soc. Sci.    1 Female      U    15
31             10        145 0.0689655            1       GRC    1 Female      U    14
32             15        120 0.1250000            1       MLL    3 Female      U    16
35             14        165 0.0848485            1      Math    5 Female      U    12
41              8        145 0.0551724            1       AGB    4 Female      U    16
42             13        135 0.0962963            1      SOCS    3 Female      U    16
44             15        165 0.0909091            1      Math    4 Female      U    12
56             17        144 0.1180560            1       Bio    2 Female      U    13
58              5        185 0.0270270            1     Psych    4   Male      U    15
63             11        180 0.0611111            1      Kine    4   Male      U    18
64              8        140 0.0571429            1  Poli Sci    3 Female      U    12
67             12        127 0.0944882            1      Phys    5 Female      U    16
70              5        125 0.0400000            1      ARCE    2 Female      U    16
72             20        180 0.1111110            1       CPE    1   Male      U    13
78              8        125 0.0640000            1      Pols    5 Female      U    16
79             15        130 0.1153850            1       Bus    3 Female      U    16
81             25        175 0.1428570            1        CE    4 Female      U    15
82             24        145 0.1655170            1       Bus    3 Female      U    16
83              5        115 0.0434783            1        ME    3 Female      U    14
85             19        130 0.1461540            1        LS    5 Female      U    16
89             20        170 0.1176470            1       Bus    3   Male      U    18
92             18        160 0.1125000            1      MFGE    3   Male      U    13
93             11        168 0.0654762            1       Bus    3   Male      U    16
# Question 2: (d)
> mean(Backpack_p$Ratio)
[1] 0.08684313
# Question 2: (e)
> backpack_f = subset(backpack, Sex == 'Female')
> backpack_f
   BackpackWeight BodyWeight     Ratio BackProblems       Major Year    Sex Status Units
1               9        125 0.0720000            1         Bio    3 Female      U    13
3              10        120 0.0833333            1         GRC    4 Female      U    14
5               8        180 0.0444444            0          EE    2 Female      U    14
8               4        185 0.0216216            1        ARCE    5 Female      U    18
9               5        130 0.0384615            0         Bio    4 Female      U    14
10              2        120 0.0166667            0         Bio    5 Female      U     8
11              8        135 0.0592593            0         Bus    3 Female      U    15
15             12        125 0.0960000            1          CS    3 Female      U    12
17             13        145 0.0896552            1         Bio    4 Female      U    16
18              6        105 0.0571429            0          LS    3 Female      U    16
19             10        125 0.0800000            1          LS    3 Female      U    17
22             14        117 0.1196580            1          EE    2 Female      U    12
24             15        205 0.0731707            0         Bio    5 Female      U    12
28              5        140 0.0357143            0    Poli Sci    2 Female      U    16
30              6        165 0.0363636            1   Soc. Sci.    1 Female      U    15
31             10        145 0.0689655            1         GRC    1 Female      U    14
32             15        120 0.1250000            1         MLL    3 Female      U    16
33              9        135 0.0666667            0          LS    6 Female      G    14
35             14        165 0.0848485            1        Math    5 Female      U    12
36             17        145 0.1172410            0       Psych    4 Female      U    13
```

```
37              14        140 0.1000000            0        ARCE    2 Female      U    15
40               8        143 0.0559441            0         Psy    2 Female      U    17
41               8        145 0.0551724            1         AGB    4 Female      U    16
42              13        135 0.0962963            1        SOCS    3 Female      U    16
43              10        130 0.0769231            0        Nut.    3 Female      U    16
44              15        165 0.0909091            1        Math    4 Female      U    12
45              10        140 0.0714286            0          LS    1 Female      U    14
46               5        115 0.0434783            0   Soc. Sci.    1 Female      U    13
47               6        128 0.0468750            0         Bus    1 Female      U    14
48               5        150 0.0333333            0         SPC    2 Female      U    14
50              10        150 0.0666667            0          CD    3 Female      U    16
53              21        116 0.1810340            0 Vocal Music    4 Female      U    16
55              13        145 0.0896552            0         Bio    4 Female      U    16
56              17        144 0.1180560            1         Bio    2 Female      U    13
57              10        130 0.0769231            0        Kine    2 Female      U    15
59               9        140 0.0642857            0        Kine    3 Female      U    16
60              13        125 0.1040000            0        Math    2 Female      U    17
62              10        150 0.0666667            0   Nutrition    4 Female      U    12
64               8        140 0.0571429            1    Poli Sci    3 Female      U    12
67              12        127 0.0944882            1        Phys    5 Female      U    16
68              14        150 0.0933333            0        Kine    3 Female      U    16
70               5        125 0.0400000            1        ARCE    2 Female      U    16
74              14        125 0.1120000            0        Pols    2 Female      U    17
76              25        144 0.1736110            0          LS    3 Female      U    17
77               2        105 0.0190476            0          IE    6 Female      U    15
78               8        125 0.0640000            1        Pols    5 Female      U    16
79              15        130 0.1153850            1         Bus    3 Female      U    16
80              10        120 0.0833333            0         AGB    4 Female      U    16
81              25        175 0.1428570            1          CE    4 Female      U    15
82              24        145 0.1655170            1         Bus    3 Female      U    16
83               5        115 0.0434783            1          ME    3 Female      U    14
84              10        110 0.0909091            0          CE    2 Female      U    16
85              19        130 0.1461540            1          LS    5 Female      U    16
86              13        135 0.0962963            0          LS    4 Female      U    16
87               9        128 0.0703125            0        Aero    1 Female      U    16
> backpack_m = subset(backpack, Sex == 'Male')
> backpack_m
    BackpackWeight BodyWeight     Ratio BackProblems      Major Year  Sex Status Units
2                8        195 0.0410256            0 Philosophy    5 Male      U    12
4                6        155 0.0387097            0        CSC    6 Male      G     0
6                5        240 0.0208333            0    History    0 Male      G     0
7                8        170 0.0470588            0         CM    3 Male      U    15
12              21        160 0.1312500            0         ME    5 Male      U    12
13              11        170 0.0647059            0        CPE    4 Male      U    16
14              11        170 0.0647059            1        CPE    4 Male      U    16
16              11        175 0.0628571            0        CPE    4 Male      U    16
20               5        165 0.0303030            0        Bio    3 Male      U    14
21              20        170 0.1176470            0        CSC    2 Male      U    16
23              11        145 0.0758621            0         CE    5 Male      U    14
25              10        270 0.0370370            0        AGB    3 Male      U    12
26              18        160 0.1125000            1       Mate    1 Male      U    13
27               8        150 0.0533333            0        AGB    4 Male      U    16
29              25        190 0.1315790            0  Aero Eng.    2 Male      U    14
34               7        145 0.0482759            0        CPE    6 Male      U    11
```

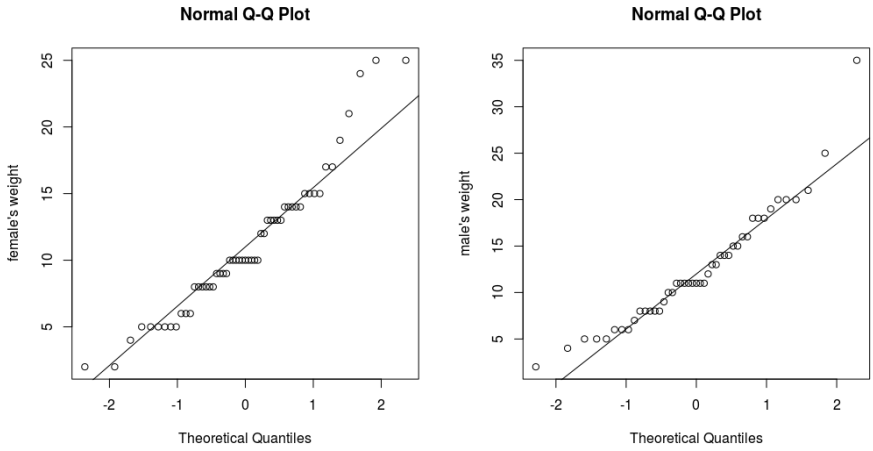
```
38              11        160 0.0687500           0         EE    4 Male      U    15
39               4        155 0.0258065           0         IT    3 Male      U    16
49              13        220 0.0590909           0         CE    2 Male      U    13
51              19        170 0.1117650           0         ME    5 Male      U    12
52               8        155 0.0516129           0       Econ    2 Male      U    12
54              16        190 0.0842105           0        Bus    2 Male      U    14
58               5        185 0.0270270           1      Psych    4 Male      U    15
61              10        175 0.0571429           0       Act.    2 Male      U    19
63              11        180 0.0611111           1       Kine    4 Male      U    18
65              12        220 0.0545455           0         EE    1 Male      U    14
66              35        195 0.1794870           0       Aero    5 Male      U    12
69               2        125 0.0160000           0         IT    5 Male      U    15
71               6        180 0.0333333           0         ME    4 Male      U    15
72              20        180 0.1111110           1        CPE    1 Male      U    13
73              14        150 0.0933333           0       Aero    2 Male      U    14
75               8        160 0.0500000           0         ME    2 Male      U    13
88              14        135 0.1037040           0       ARCE    4 Male      U    13
89              20        170 0.1176470           1        Bus    3 Male      U    18
90              16        175 0.0914286           0        Bio    3 Male      U    12
91              18        150 0.1200000           0         IT    6 Male      U    17
92              18        160 0.1125000           1       MFGE    3 Male      U    13
93              11        168 0.0654762           1        Bus    3 Male      U    16
94              13        155 0.0838710           0        Psy    4 Male      U    14
95              15        210 0.0714286           0       Arch    3 Male      U    17
96              14        165 0.0848485           0         ME    3 Male      U    11
97               6        195 0.0307692           0       APIO    1 Male      U    16
98              11        130 0.0846154           0        Bus    1 Male      U    12
99               9        140 0.0642857           0       AERO    3 Male      U    12
100             15        170 0.0882353           0    History    5 Male      U    14
```

```
# Four steps of statistical  modeling
# Step 1: Choose a model
> par(mfrow = c(1,1))
> boxplot(backpack_f$BackpackWeight, backpack_m$BackpackWeight, main = "Backpack weight", names
= c("females", "males"))
```

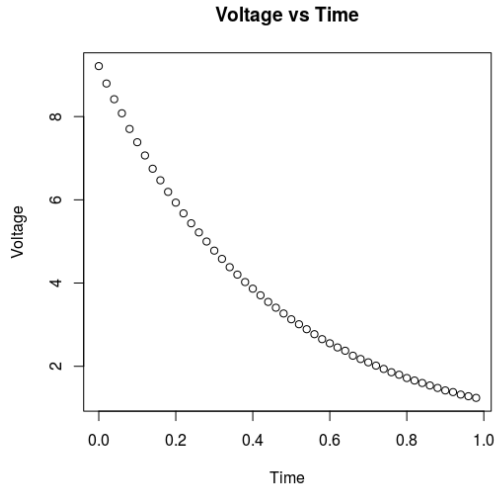



```
# Step 2: Fit the model
# First find if the variances of these two dataset are same or not, then using two sample t
```

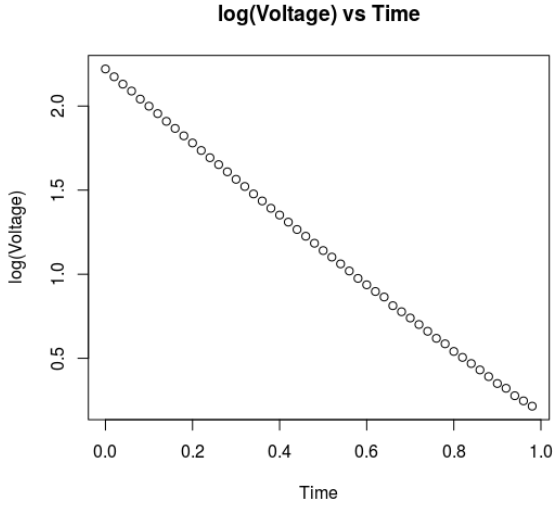
```
test to fit the model. For there the variances are different, thus, the var.equal should be set
up to false, which is default.
> var(backpack_f$BackpackWeight)
[1] 27.96162
> var(backpack_m$BackpackWeight)
[1] 39.38586
> t.test(backpack_f$BackpackWeight, backpack_m$BackpackWeight)

        Welch Two Sample t-test

data:  backpack_f$BackpackWeight and backpack_m$BackpackWeight
t = -1.1782, df = 86.25, p-value = 0.242
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.7241290  0.9524118
sample estimates:
mean of x mean of y
 11.03636  12.42222

# Step 3: Assess the model
> length(backpack_m$BackpackWeight)
[1] 45
> length(backpack_f$BackpackWeight)
[1] 55
> par(mfrow = c(1,2))
> qqnorm(backpack_f$BackpackWeight, ylab = "female's weight")
> qqline(backpack_f$BackpackWeight, ylab = "female's weight")
> qqnorm(backpack_m$BackpackWeight, ylab = "male's weight")
> qqline(backpack_m$BackpackWeight, ylab = "male's weight")
```




```
# Step 4: Use the model
The two subsets of this model are not very large. Since the p_value is 0.242 which is very
larger than 0.05, the assumption is not correct and then sample is not that unusual. Thus,
there is difference of backpack weight between male and female students.


> par(mfrow = c(1,1))

## Question 3
# Question 3: (a)
> data(Volts)
> plot(Volts$Time, Volts$Voltage, main = "Voltage vs Time", xlab = "Time", ylab = "Voltage")
```

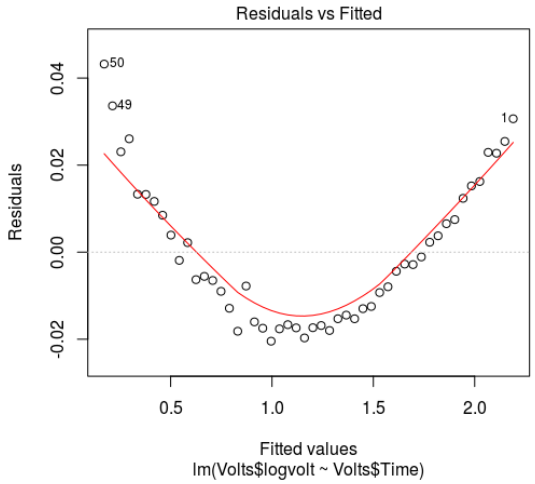



```
# Comment for (a): the plot shows that as the time increasing, the voltage will get lower and lower,
and the pattern is non-linear.
##############################################
# Question 3: (b)
> Volts = transform(Volts, logvolt = log(Voltage))
> plot(Volts$Time, Volts$logvolt, main = "log(Voltage) vs Time", xlab = "Time", ylab = "log(Voltage)")
```

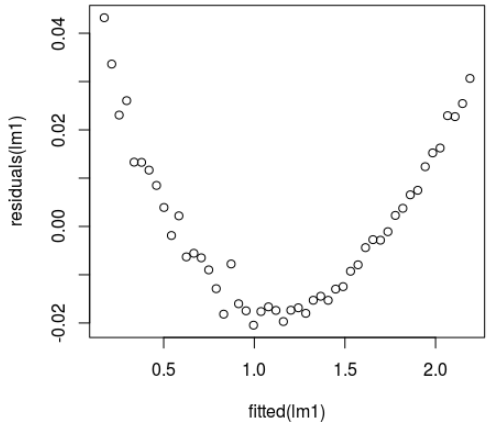



```
# Comment for (b): the plot shows that after transform the voltage into log(voltage), the plot attern
is more like linear.
##############################################
# Question 3: (c)
> lm1 = lm(Volts$logvolt ~ Volts$Time)
> summary(lm1)

Call:
lm(formula = Volts$logvolt ~ Volts$Time)

Residuals:
       Min         1Q     Median         3Q        Max
-0.020448 -0.015084 -0.003621  0.012190  0.043212
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.189945   0.004637   472.3   <2e-16 ***
Volts$Time  -2.059065   0.008154  -252.5   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01664 on 48 degrees of freedom
Multiple R-squared:  0.9992,    Adjusted R-squared:  0.9992
F-statistic: 6.377e+04 on 1 and 48 DF,  p-value: < 2.2e-16

# Comment for (c): the prediction equation is hat(logVoltage) = 2.189 - 2.059*Time.
################################################
# Question 3: (d)
> plot(lm1)
Hit <Return> to see next plot:
Hit <Return> to see next plot: plot(lm1)
Hit <Return> to see next plot:
Hit <Return> to see next plot:
```




```
> plot(fitted(lm1), residuals(lm1))
```



```
# Comment for (d): the plot shows that it is a curved pattern. Since the logVoltage plot shows the
linear pattern, this means that using this model will make the prediction that might be higher than
real in the middle and lower at the boundaries of time.
```