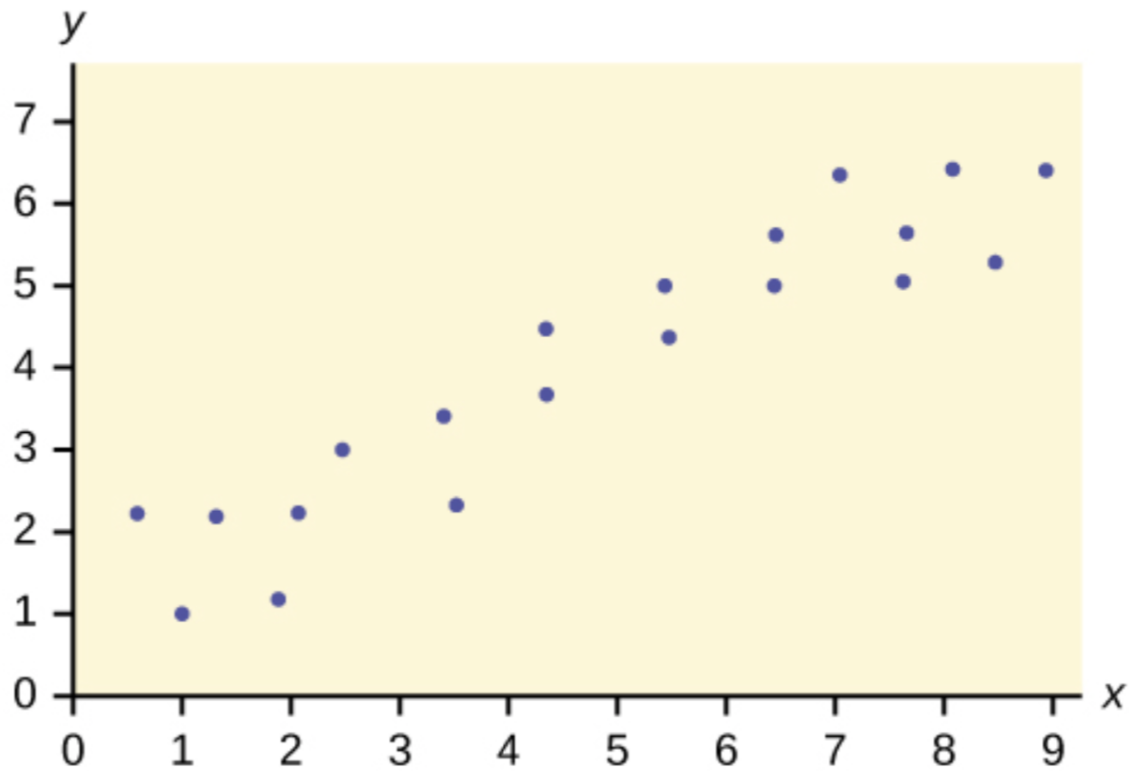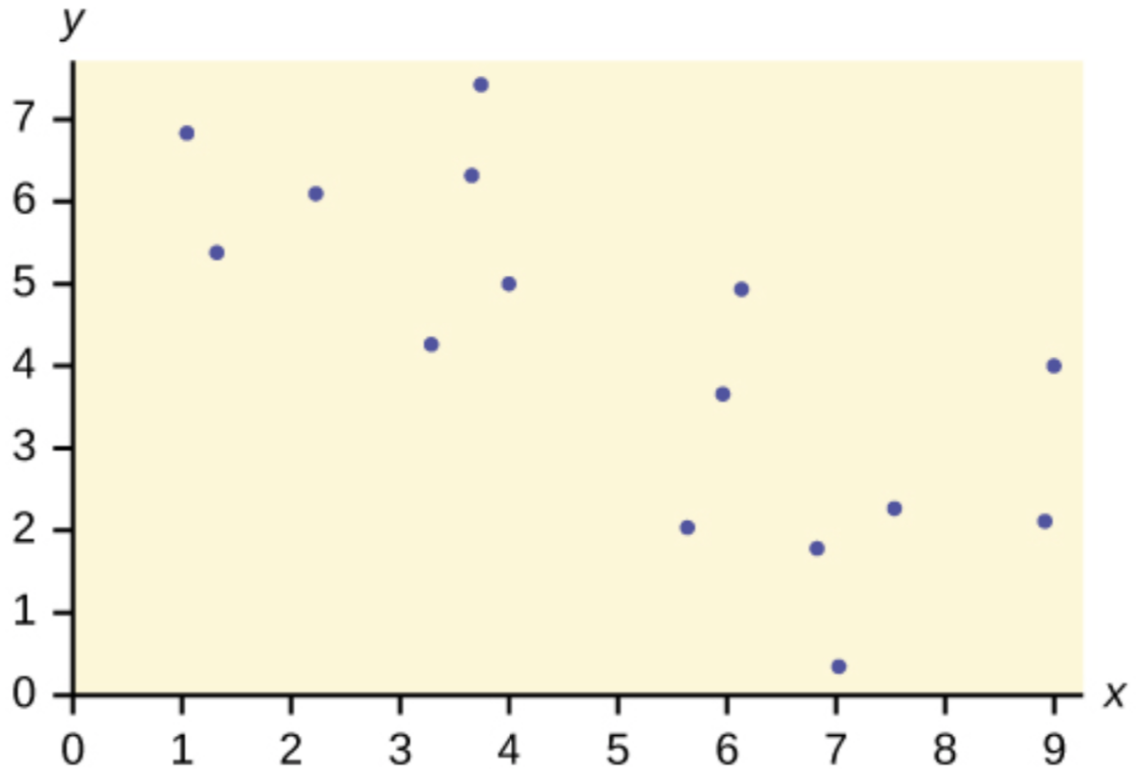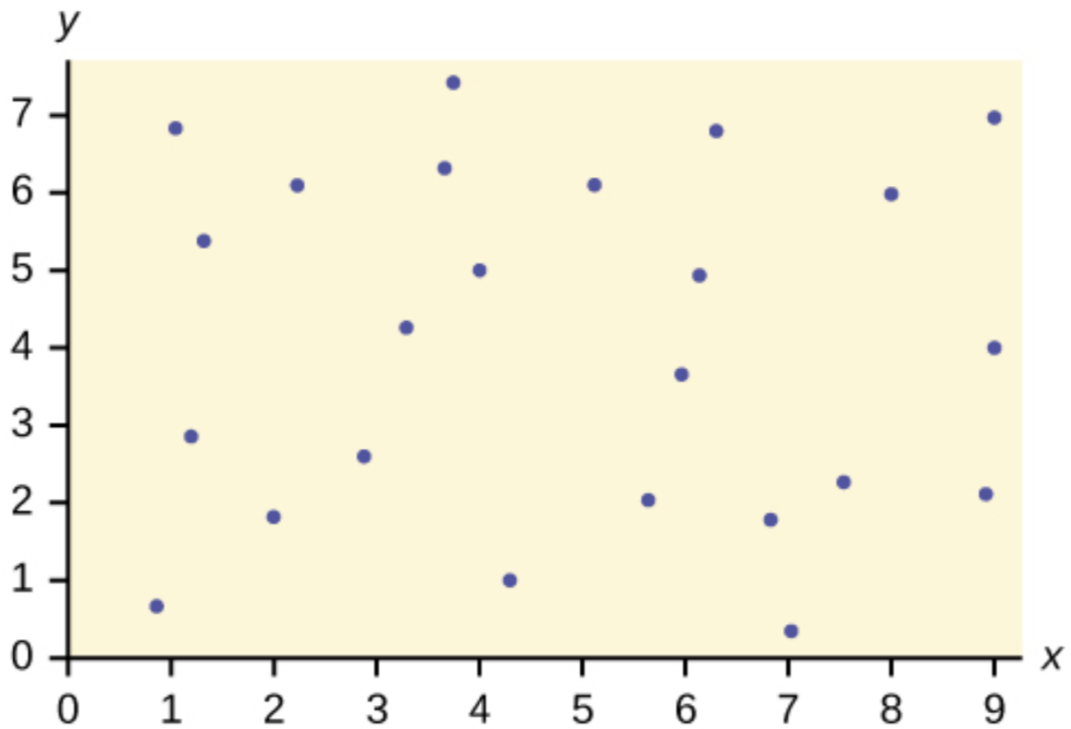# Exercise set 6
Stat 251

(1) Does the scatterplot appear to be linear? Is it positive or negative? Weak or strong?

(2) Does the scatterplot appear to be linear? Is it positive or negative? Weak or strong?

(3) Does the scatterplot appear to be linear? Is it positive or negative? Weak or strong?
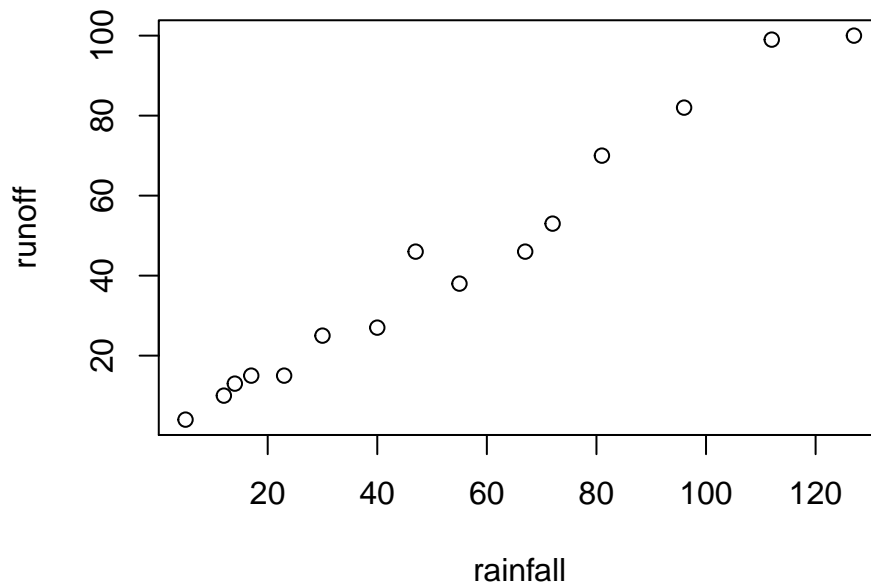
(4) The article "Characterization of Highway Runoff in Austin, Texas, Area"[1] gave a scatterplot, along with the regression equation that modeled runoff by rainfall (with units of measurements volumes of water in m^3) for a particular location. All output needed for this is included below.

(a) State the population model; identify all components of the model
(b) Does the relationship between rainfall volume and runoff volume appear to be linear? Describe the scatterplot
(c) State the regression equation (with the values from the regression output)
(d) Interpret the slope and intercept in context. If one does not make logical sense, state why not.
(e) Using the regression equation, estimate the runoff volume for rainfall amounts of 5, 47, and 127 m^3
(f) Calculate the residuals for each estimated value. The observed values of y for $x = 5, \ 47, \ 127$ are (respectively): $y = 4, \ 46, \ 100$
(g) Conduct a test of significance for the slope (results from R output). List hypotheses, test statistic, pvalue, result, and conclusion.
(h) List the values of r and R^2 (r will have to be computed). Interpret both values in context. What do they tell us about the model?
(i) List assumptions of regression. Check the assumptions (briefly describe how each assumption is met or not).
(j) Overall, what is your assessment of the model? Good...bad? Use results from significance test, r, R^2, and assumptions to make this assessment.

```
head(texas)
```

```
  rainfall runoff
1        5      4
2       12     10
3       14     13
4       17     15
5       23     15
6       30     25
```

```
with(texas,plot(runoff~rainfall,main='Raw Data Scatterplot'))
```

Raw Data Scatterplot

[1]"Characterization of Highway Runoff in Austin, Texas Area", *J. of Envir. Engr.*, 1998: 131-137

```
sandy=lm(runoff~rainfall,data=texas)
summary(sandy)

Call:
lm(formula = runoff ~ rainfall, data = texas)

Residuals:
   Min     1Q Median     3Q    Max
-8.279 -4.424  1.205  3.145  8.261

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.12830    2.36778  -0.477    0.642
rainfall     0.82697    0.03652  22.642  7.9e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.24 on 13 degrees of freedom
Multiple R-squared:  0.9753,    Adjusted R-squared:  0.9734
F-statistic: 512.7 on 1 and 13 DF,  p-value: 7.896e-12
```
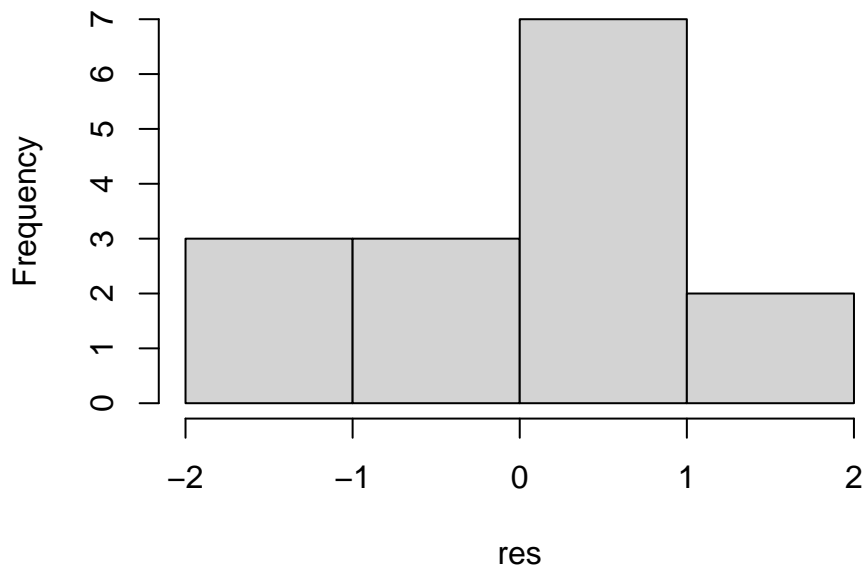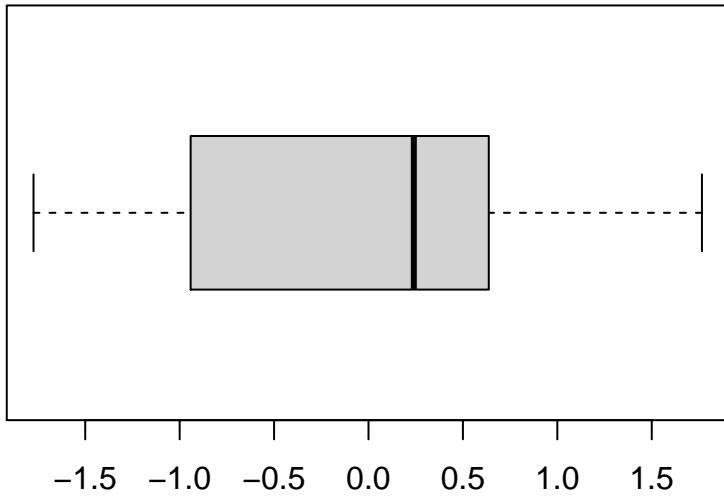
```
res=rstudent(sandy); pred=fitted(sandy)
hist(res,main='Histogram of residuals')
```
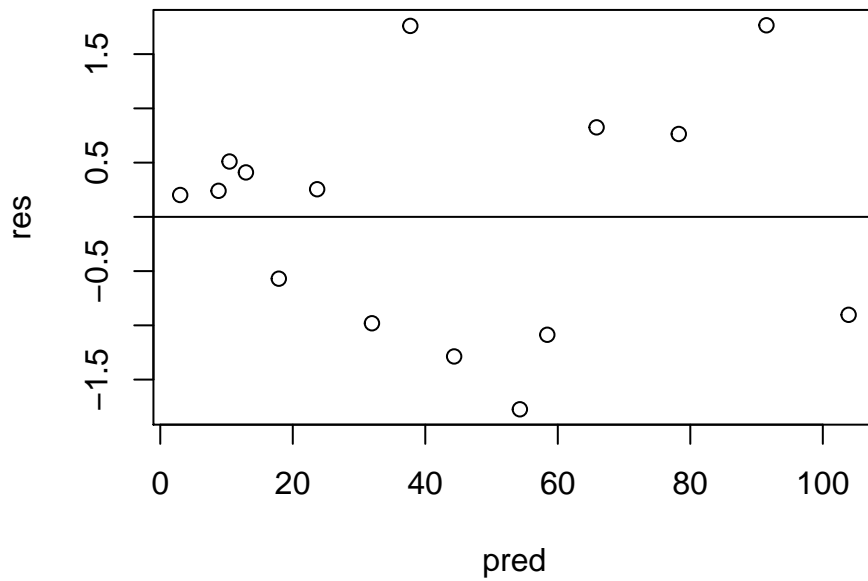
Histogram of residuals


```
boxplot(res,horizontal=T,main='Residuals')
```

Residuals


```
plot(pred,res,main='Residuals Vs. Predicted'); abline(0,0)
```

Residuals Vs. Predicted


```
qqnorm(res); qqline(res)
```

Normal Q-Q Plot