1a.

```
xy <- read.table("data/az-5000.txt", header = TRUE)
head(xy, 1)
tail(xy, 1)
```

1b & 1c.

```
set.seed(101) # Set Seed so that same sample can be reproduced in future also
# Now Selecting 75% of data as sample from total 'n' rows of the data
sample <- sample.int(n = nrow(xy), size = floor(.8*nrow(xy)), replace = F)
# train <- data[sample, ]
# test  <- data[-sample, ]
table(sample)
```

3.

```
#loading credit data
credit.xy <- read.table("data/credit_data.txt", header = TRUE)
dim(credit.xy)

set.seed(123)

#selecting 80% of data as training data
train <- sample(1:885, 0.8*885)

#showing the number of cases per class for both training and test data
table(credit.xy$Fail[train])
table(credit.xy$Fail[-train])

#(a)
#fitting logistic regression to training data and creating summary data
credit.glm <- glm(Fail~.-Id, family = binomial, data = credit.xy[train,])
summary(credit.glm)
#(c)
#predicting if a firm will go bankrupt if the predicted probability is
#P(Y=1|X=x) of bankruptcy is 0.5 or greater and creating a confusion matrix

yHat <- predict(credit.glm , credit.xy[-train, c(1,3:15)], type = "response")
table(credit.xy$Fail[-train], yHat >= 0.5)
```

6a.

```
set.seed(1)
y = rnorm(100)
x = rnorm(100)
y = x - 2*x^2 + rnorm(100)
```

```
6b.
plot(x, y)
6c.
#i.
library(boot)
set.seed(1)
Data <- data.frame(x, y)
fit.glm.1 <- glm(y ~ x)
cv.glm(Data, fit.glm.1)$delta[1]
#ii.
fit.glm.2 <- glm(y ~ poly(x, 2))
cv.glm(Data, fit.glm.2)$delta[1]
#iii.
fit.glm.3 <- glm(y ~ poly(x, 3))
cv.glm(Data, fit.glm.3)$delta[1]
#iv.
fit.glm.4 <- glm(y ~ poly(x, 4))
cv.glm(Data, fit.glm.4)$delta[1]
6d.
#i.
set.seed(10)
fit.glm.1 <- glm(y ~ x)
cv.glm(Data, fit.glm.1)$delta[1]
#ii.
fit.glm.2 <- glm(y ~ poly(x, 2))
cv.glm(Data, fit.glm.2)$delta[1]
#iii.
fit.glm.3 <- glm(y ~ poly(x, 3))
cv.glm(Data, fit.glm.3)$delta[1]
#iv.
fit.glm.4 <- glm(y ~ poly(x, 4))
cv.glm(Data, fit.glm.4)$delta[1]
6f.
summary(fit.glm.4)
```