5.

- a. Verify that SSPE = 3.036669

```
lof <- data.frame(x = c(1,1,2,3.3,3.3,4,4,4,4.7,5,5.6,5.6,5.6,6,6,6.5,6.9),
                  y = c(2.3,1.8,2.8,1.8,3.7,2.6,2.6,2.2,3.2,2,3.5,2.8,2.1,3.4,3.2,3.4,5))
```

$$SSPE = \sum_j \sum_i (Y_{ij} - \bar{Y_j})^2$$

```
# calculate group mean
mean_y <- lof %>%
  group_by(x) %>%
  summarise(meany = mean(y), n = n())
# calculate SSPE
(lof %>%
  left_join(mean_y, by = c("x" = "x")) %>%
  mutate(err2 = (y - meany)^2) %>%
  summarise(sum(err2)))
```

```
##   sum(err2)
## 1  3.036667
```

- b. Would a cubic polynomial fit well for the same data set? What about a quintic polynomial?

```
lof <- lof %>%
  mutate(x2 = x^2) %>%
  mutate(x3 = x^3) %>%
  mutate(x4 = x^4) %>%
  mutate(x5 = x^5)
```

```
mod1 <- lm(y ~ x, data = lof)
mod3 <- lm(y ~ x + x2 + x3, data = lof)
mod5 <- lm(y ~ x + x2 + x3 + x4 +x5, data = lof)
```

```
anova(mod1, mod3)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x
## Model 2: y ~ x + x2 + x3
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     15 7.4150
## 2     13 4.2043  2    3.2107 4.9639 0.02502 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(mod1, mod5)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x
## Model 2: y ~ x + x2 + x3 + x4 + x5
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     15 7.4150
## 2     11 4.1306  4    3.2843 2.1866 0.1376
```

```r
anova(mod3, mod5)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x + x2 + x3
## Model 2: y ~ x + x2 + x3 + x4 + x5
##   Res.Df    RSS Df Sum of Sq     F Pr(>F)
## 1     13 4.2043
## 2     11 4.1306  2  0.073615 0.098 0.9074
```

The cubic polynomial has a significant better fit than the model with only linear terms. While for the quintic polynomial model, it has neither significantly better fit than the linear model nor than the cubic polynomial model.