# Notes: MS 204 Chapter 5.2-5.3

### Overview

- Outliers
- Inference

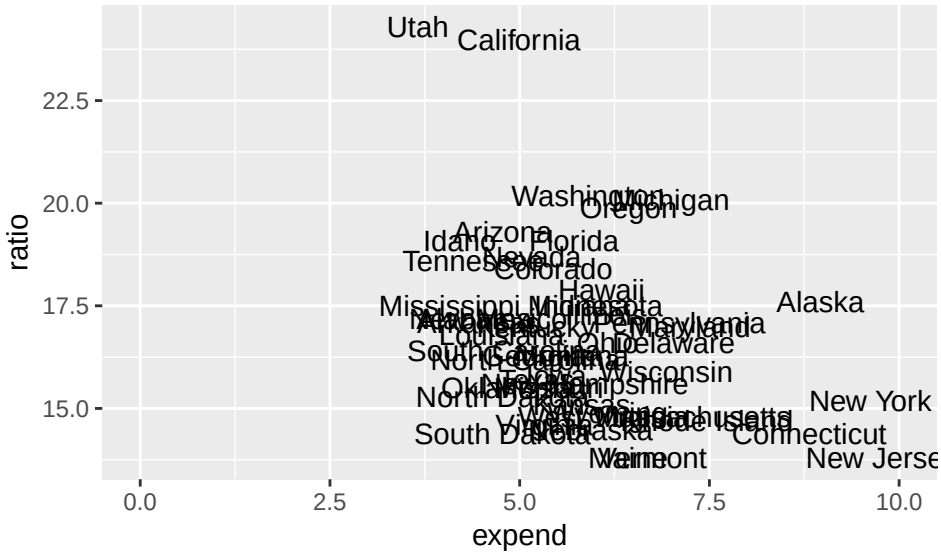
## Outliers and regression

Ex: $X$ = expend, $Y$ = ratio

```
library(tidyverse); library(mosaic)
SAT %>% summarise(cor.SAT = cor(expend, ratio))
```

```
##      cor.SAT
## 1 -0.3710254
```

```
qplot(x = expend, y = ratio, data = SAT, geom = "text", label = state) + xlim(c(0, 10))
```




```
SAT %>% filter(ratio < 21) %>% summarise(cor.SAT = cor(expend, ratio))
```

```
##      cor.SAT
## 1 -0.2919665
```

```
fit <- lm(ratio ~ expend, data = SAT)
msummary(fit)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20.5016     1.3502  15.184  < 2e-16 ***
## expend       -0.6170     0.2229  -2.768  0.00799 **
##
## Residual standard error: 2.126 on 48 degrees of freedom
```

```
## Multiple R-squared:  0.1377, Adjusted R-squared:  0.1197
## F-statistic: 7.662 on 1 and 48 DF,  p-value: 0.007987
```

```
SAT.filter <- SAT %>% filter(ratio < 21)
fit <- lm(ratio ~ expend, data = SAT.filter)
msummary(fit)
```

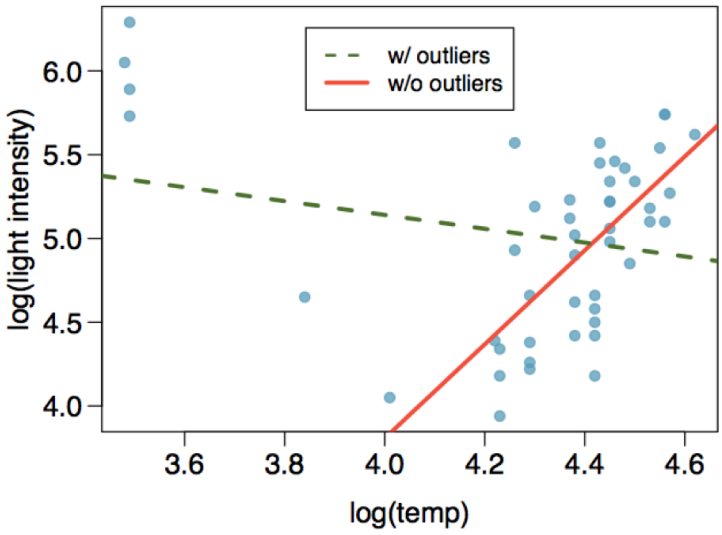
```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18.7997     1.1112   16.92   <2e-16 ***
## expend       -0.3761     0.1816   -2.07   0.0441 *
##
## Residual standard error: 1.674 on 46 degrees of freedom
## Multiple R-squared:  0.08524,    Adjusted R-squared:  0.06536
## F-statistic: 4.287 on 1 and 46 DF,  p-value: 0.04405
```

**Outliers**

**High leverage points**

**Influential points**

**Example: log of the surface temperature and the log of the light intensity of 47 stars in the star cluster CYG OB1**



### Inference & confidence intervals

```
fit <- lm(ratio ~ expend, data = SAT)
msummary(fit)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20.5016     1.3502  15.184  < 2e-16 ***
## expend       -0.6170     0.2229  -2.768  0.00799 **
##
## Residual standard error: 2.126 on 48 degrees of freedom
## Multiple R-squared:  0.1377, Adjusted R-squared:  0.1197
## F-statistic: 7.662 on 1 and 48 DF,  p-value: 0.007987
```

### Confidence intervals

```
confint(fit, level = 0.95)
```

```
##                2.5 %     97.5 %
## (Intercept) 17.78688 23.2164128
## expend      -1.06519 -0.1688444
```