newspapers

Chris Parrish

December 8, 2016

newspapers

reference:
- Sheather, A Modern Approach to Regression with R, pp.5-7

Load packages.

```
library(ggplot2)
library(dplyr)
```

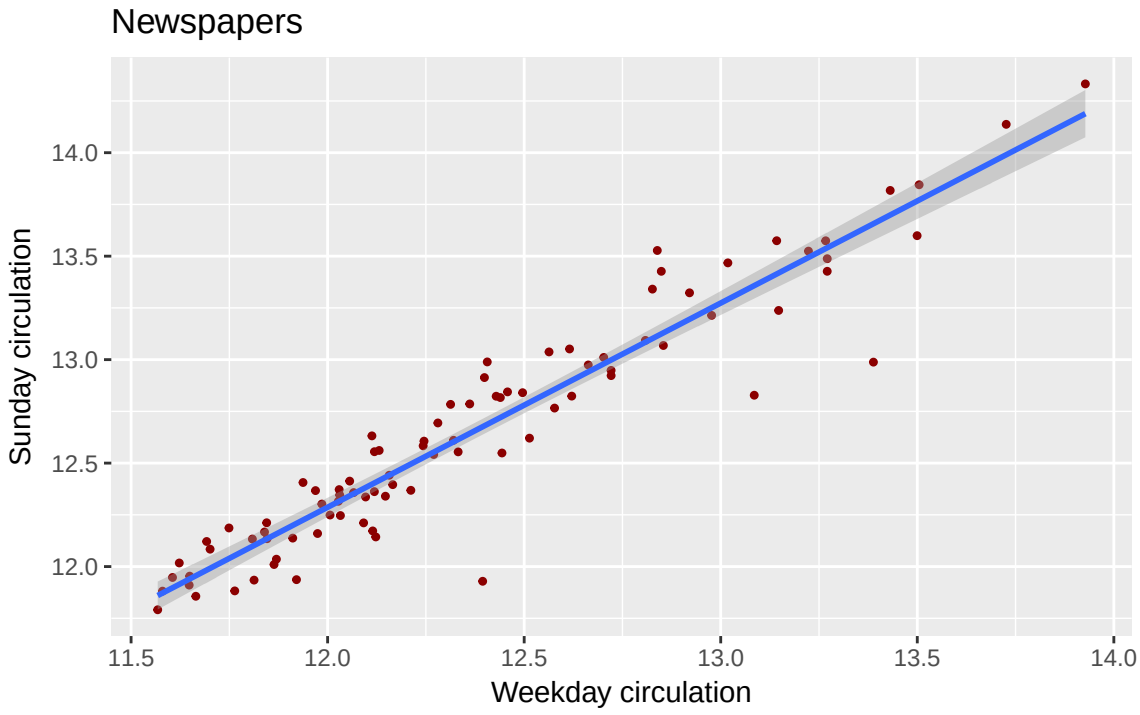
Import the data.

```
data <- read.delim("circulation.txt", header=TRUE)
colnames(data) <- c("Newspaper", "Sunday", "Weekday", "Tabloid")
head(data, 3)
```

```
##                       Newspaper Sunday Weekday Tabloid
## 1   Akron (OH) Beacon Journal  185915  134401       0
## 2     Albuquerque (NM) Journal 154413  109693       0
## 3 Allentown (PA) Morning Call  165607  111594       0
```

Visualize the data.

```
ggplot(data, aes(log(Weekday), log(Sunday))) +
  geom_point(shape = 20, color = "darkred") +
  geom_smooth(method = "lm") +
  labs(x = "Weekday circulation", y = "Sunday circulation", title = "Newspapers")
```



Linear model.

$$log(Sunday) \sim log(Weekday)$$

```
newspaper.lm <- lm(log(Sunday) ~ log(Weekday), data=data)
summary(newspaper.lm)
```

```
##
## Call:
## lm(formula = log(Sunday) ~ log(Weekday), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74689 -0.07547  0.01328  0.11339  0.41339
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.44511    0.43204    1.03    0.306
## log(Weekday)  0.98679    0.03488   28.29   <2e-16
##
## Residual standard error: 0.181 on 87 degrees of freedom
## Multiple R-squared:  0.9019, Adjusted R-squared:  0.9008
## F-statistic: 800.2 on 1 and 87 DF,  p-value: < 2.2e-16
```

Identify tabloids

```
data$Type <- factor(data$Tabloid, labels = c("Serious", "Tabloid"))
ggplot(data, aes(log(Weekday), log(Sunday))) +
  geom_point(aes(color = Type), shape = 20) +
  labs(x = "Weekday circulation", y = "Sunday circulation", title = "Newspapers")
```