

SDM4 in R: Scatterplots, Association, and Correlation (Chapter 6)

Nicholas Horton (nhorton@amherst.edu)

August 12, 2017

Introduction and background

This document is intended to help describe how to undertake analyses introduced as examples in the Fourth Edition of *Stats: Data and Models* (2014) by De Veaux, Velleman, and Bock. More information about the book can be found at http://wps.aw.com/aw_deveaux_stats_series. This file as well as the associated R Markdown reproducible analysis source file used to create it can be found at <http://nhorton.people.amherst.edu/sdm4>.

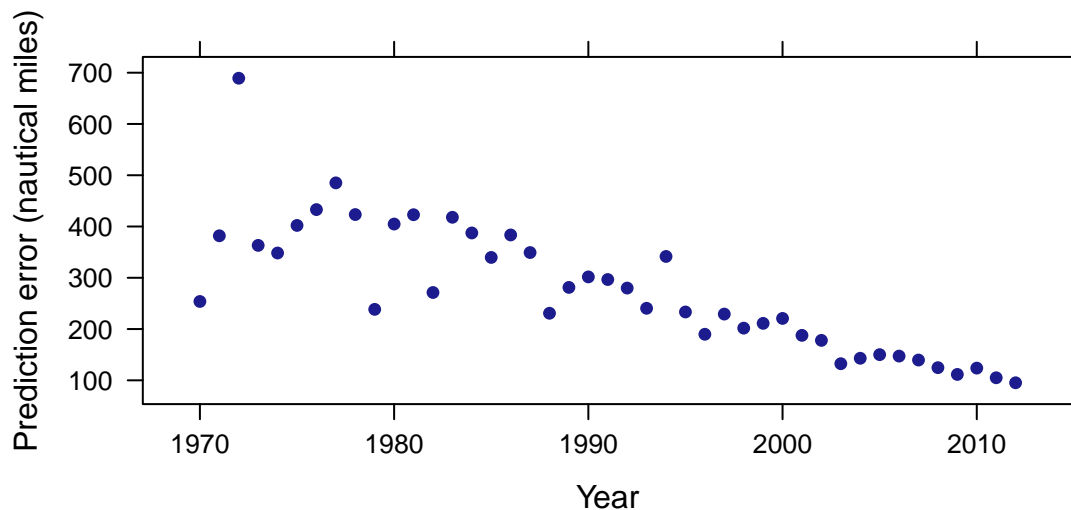
This work leverages initiatives undertaken by Project MOSAIC (<http://www.mosaic-web.org>), an NSF-funded effort to improve the teaching of statistics, calculus, science and computing in the undergraduate curriculum. In particular, we utilize the `mosaic` package, which was written to simplify the use of R for introductory statistics courses. A short summary of the R needed to teach introductory statistics can be found in the `mosaic` package vignettes (<http://cran.r-project.org/web/packages/mosaic>). A paper describing the `mosaic` approach was published in the *R Journal*: <https://journal.r-project.org/archive/2017/RJ-2017-024>.

Chapter 6: Scatterplots, Association, and Correlation

Section 6.1: Scatterplots

Figure 6.1 (page 152) displays the scatterplot of the average tracking error over time.

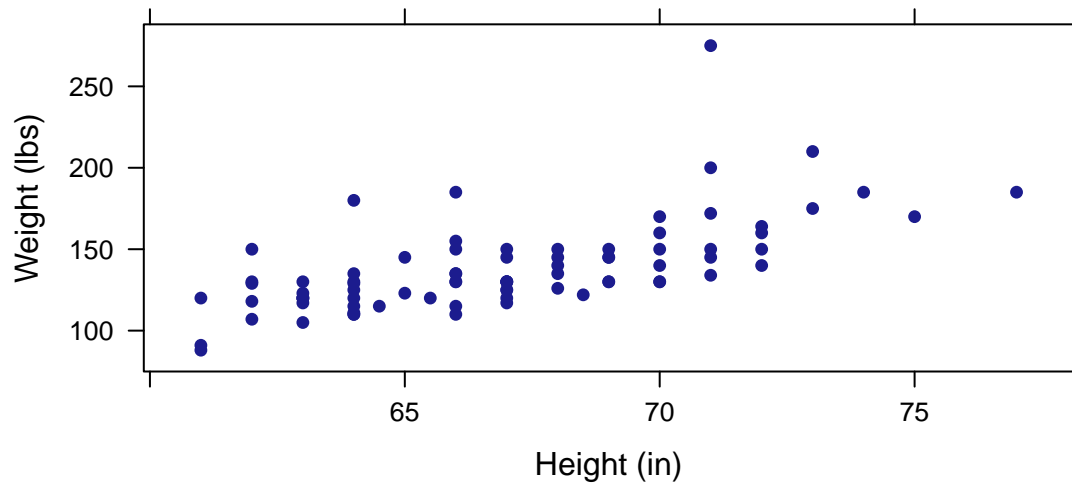
```
library(mosaic); library(readr)
options(digits=3)
Hurricanes <-
  read_csv("http://nhorton.people.amherst.edu/sdm4/data/Tracking_hurricanes_2012.csv")
xyplot(Error72h ~ Year, ylab="Prediction error (nautical miles)", data=Hurricanes)
```



Section 6.2: Correlation

Figure 6.2 (page 155) displays the scatterplot of weight vs. height for a sample of students from statistics classes.

```
HtWt <- read_csv("http://nhorton.people.amherst.edu/sdm4/data/Heights_and_Weights.csv")
xyplot(Weight ~ Height, ylab="Weight (lbs)", xlab="Height (in)", data=HtWt)
```



```
cor(Weight ~ Height, data=HtWt)
```

```
## [1] 0.644
```

Kendall's Tau and Spearman's Rho

```
cor(Weight ~ Height, method="kendall", data=HtWt)
```

```
## [1] 0.545
```

```
cor(Weight ~ Height, method="spearman", data=HtWt)
```

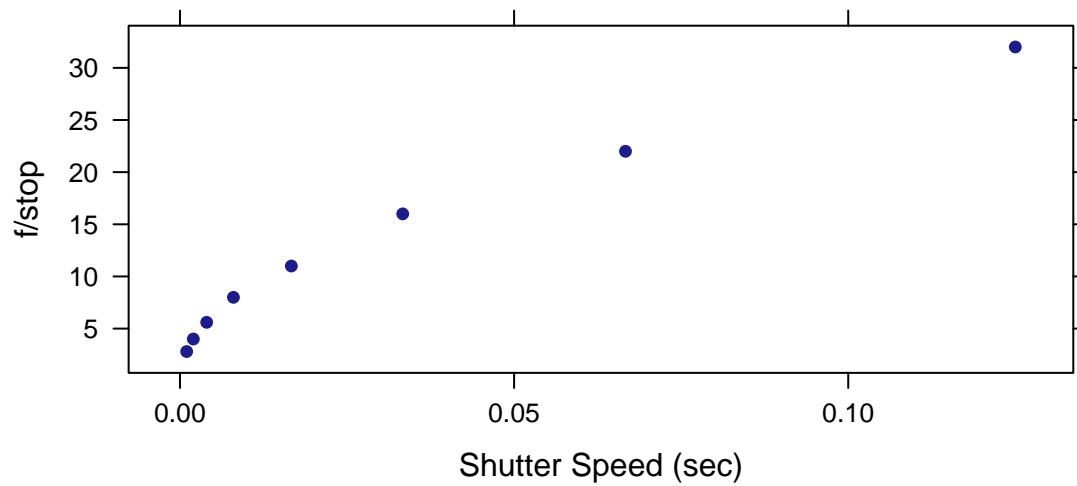
```
## [1] 0.697
```

Section 6.3: Warning: Correlation does not always equal Causation

Section 6.4: Straightening scatterplots

Since the dataset is so small for Figure 6.10 (page 165) we can enter it by hand.

```
fstop <- c(2.8, 4, 5.6, 8, 11, 16, 22, 32)
shutter <- c(1/1000, 1/500, 1/250, 1/125, 1/60, 1/30, 1/15, 1/8)
lenses <- data.frame(fstop, shutter)
xyplot(fstop ~ shutter, ylab="f/stop", xlab="Shutter Speed (sec)", data=lenses)
```



A new transformed variable can be added using the `mutate` function.

```
lenses <- mutate(lenses, fstopsq = fstop*fstop)
xyplot(fstopsq ~ shutter, ylab="f/stop (squared)", xlab="Shutter Speed (sec)", data=lenses)
```

