A Brief Introduction to Reproducible Research in RStudio with **knitr**: Examples

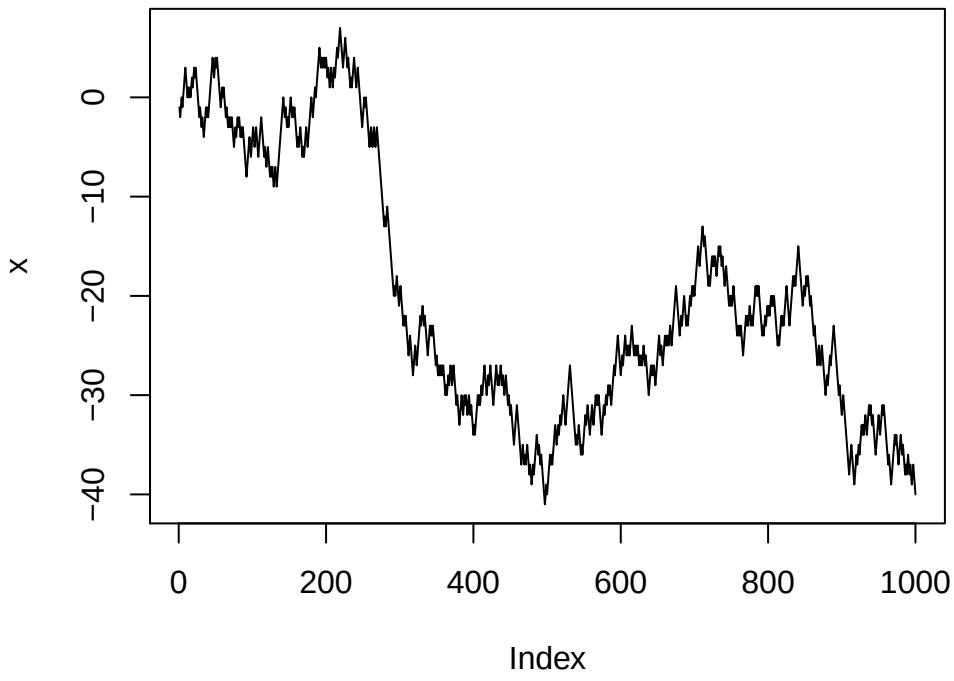
Will Massengill

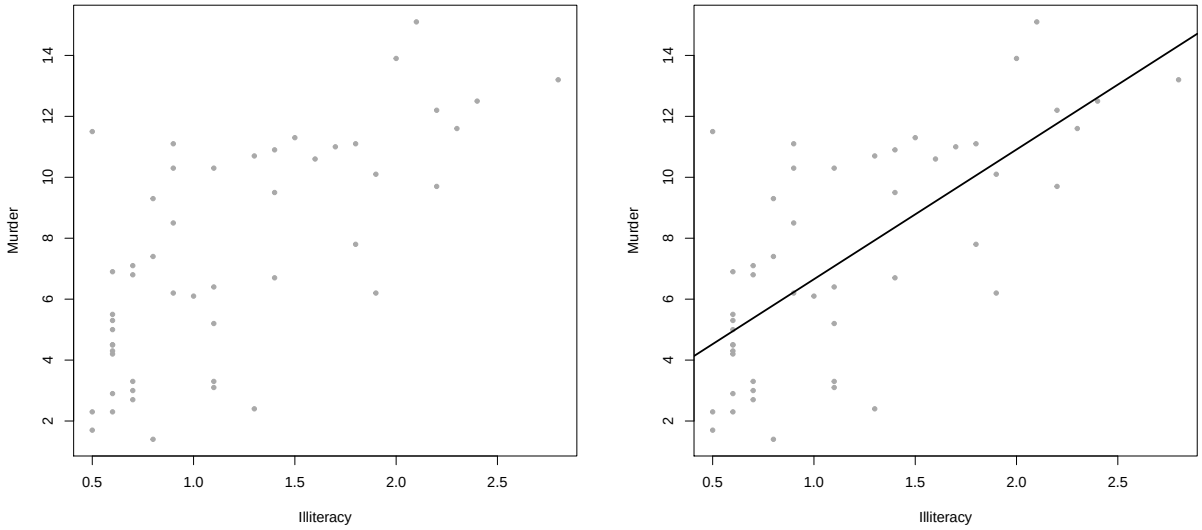
October 28, 2015

1. The first example draws a sample and demonstrates the `diff` function.

```
set.seed(50)
x <- sample.int(10, 7, replace = TRUE)
x; diff(x)

## [1] 8 5 3 8 6 1 7
## [1] -3 -2  5 -2 -5  6

identical(diff(x), x[-1] - x[-length(x)])

## [1] TRUE
```

2.
```
set.seed(1)
n <- 1000
x <- cumsum(sample(c(-1, 1), n, TRUE))
par(mar = c(4, 4, 0.1, 0.3))
plot(x, type = "l")
```





(a) Points Only

(b) The Regression

Figure 1: Illiteracy and Murder Rate

3. We can create a couple plots with the following code.

```
attach(data.frame(state.x77))
fit <- lm(Murder ~ Illiteracy)
plot(Illiteracy, Murder,  pch = 20, col = 'darkgrey')
abline(fit, lwd = 2)
```

   And we can discuss the results. As Figure 1 indicates, the slope of the regression is 4.26.

4. We can also discuss the results using a code chunk.

```
cat("The linear regression", sprintf("$Murder = %.02f + %.02f Illiteracy$...",
                                     coef(fit)[1], coef(fit)[2]))
```

   The linear regression $Murder = 2.40 + 4.26 Illiteracy$...

Table 1: Regressions of Murder Rate

	Model 1	Model 2	Model 3
(Intercept)	2.40**	−17.94	−29.40
	(0.82)	(26.42)	(24.60)
Illiteracy	4.26***	4.52***	4.53***
	(0.62)	(0.71)	(0.66)
log(Income)		2.39	2.58
		(3.10)	(2.85)
log(Area)			0.92**
			(0.30)
N	50	50	50
R^2	0.49	0.50	0.59
adj. R^2	0.48	0.48	0.56
Resid. sd	2.65	2.66	2.45

Standard errors in parentheses

† significant at $p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$

5. Let's estimate some models so that we can print a table.

```
library(apsrtable)
fit2 <- lm(Murder ~ Illiteracy + log(Income))
fit3 <- lm(Murder ~ Illiteracy + log(Income) + log(Area))
```