PLSC 503 – Spring 2018
Stupid Regression Tricks

February 6, 2018

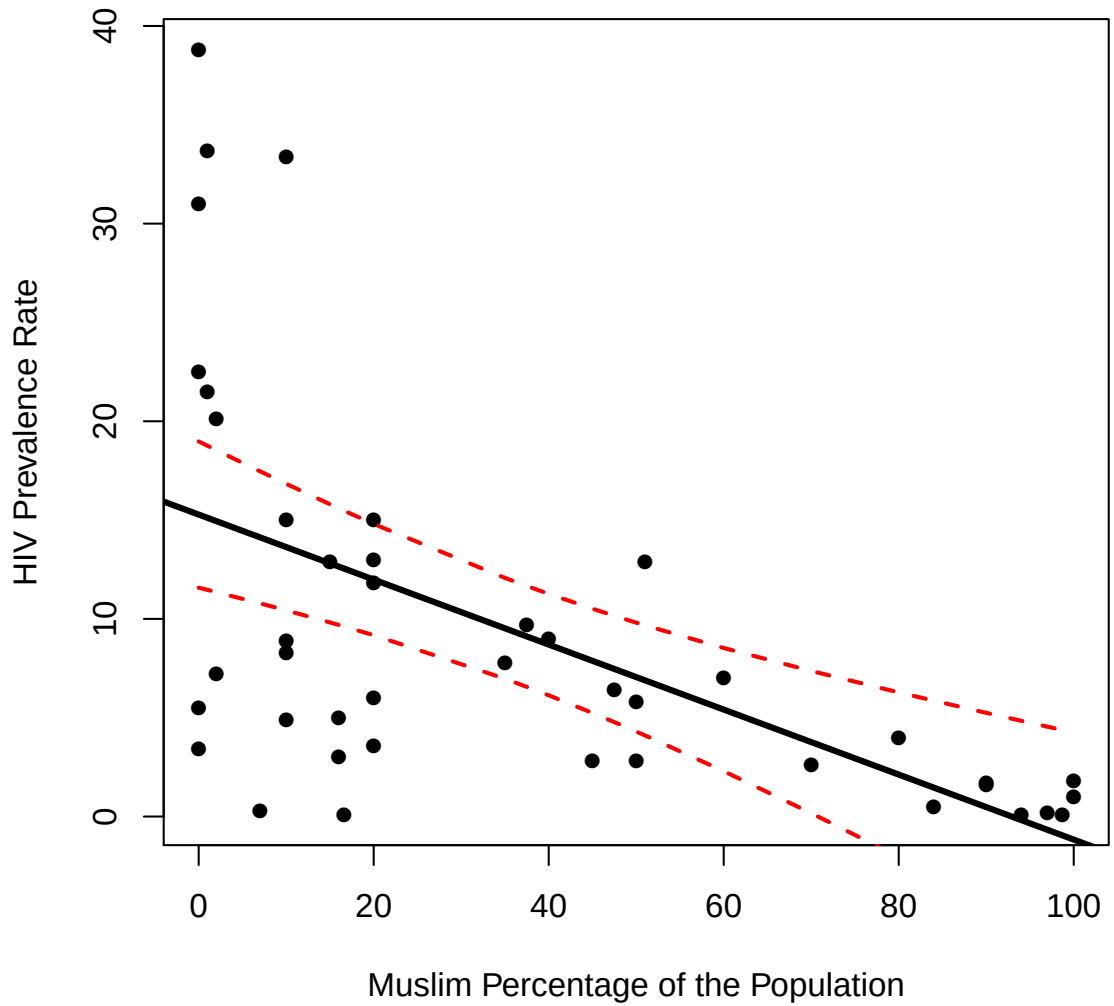
Africa (2001) Data

```
> temp<-getURL("https://raw.githubusercontent.com/PrisonRodeo/
  PLSC503-Spring-2018-git/master/Data/africa2001.csv")
> africa<-read.csv(text=temp, header=TRUE)
> summary(africa)
     ccode          cabbr                         country     population           popthou
 Min.   :404   AGO    : 1   Angola                   : 1   Min.   :   470000   Min.   :   470
 1st Qu.:452   BDI    : 1   Benin                    : 1   1st Qu.:  3446000   1st Qu.:  3446
 Median :510   BEN    : 1   Botswana                 : 1   Median :  9662000   Median :  9662
 Mean   :510   BWA    : 1   Burundi                  : 1   Mean   : 17388558   Mean   : 17390
 3rd Qu.:556   CAF    : 1   Cameroon                 : 1   3rd Qu.: 19150000   3rd Qu.: 19189
 Max.   :651   CIV    : 1   Central African Republic: 1   Max.   :117000000   Max.   :116929
               (Other):37   (Other)                  :37
     popden            polity          gdppppd          tradegdp           war             adrate
 Min.   :0.0022   Min.   :-9.000   Min.   : 0.500   Min.   :  4.03   Min.   :0.000   Min.   : 0.10
 1st Qu.:0.0134   1st Qu.:-4.500   1st Qu.: 0.855   1st Qu.:  7.64   1st Qu.:0.000   1st Qu.: 2.70
 Median :0.0357   Median : 0.000   Median : 1.200   Median : 13.56   Median :0.000   Median : 6.00
 Mean   :0.0643   Mean   : 0.512   Mean   : 2.159   Mean   : 30.49   Mean   :0.116   Mean   : 9.37
 3rd Qu.:0.0683   3rd Qu.: 5.500   3rd Qu.: 2.040   3rd Qu.: 30.01   3rd Qu.:0.000   3rd Qu.:12.90
 Max.   :0.5740   Max.   :10.000   Max.   :10.800   Max.   :272.69   Max.   :1.000   Max.   :38.80

   healthexp              subsaharan    muslperc        literacy     internalwar      intensity
 Min.   :2.00   Not Sub-Saharan: 6   Min.   :  0.0   Min.   :17.0   Min.   :0.000   Min.   :0.000
 1st Qu.:3.45   Sub-Saharan    :37   1st Qu.: 10.0   1st Qu.:43.0   1st Qu.:0.000   1st Qu.:0.000
 Median :4.40                       Median : 20.0   Median :61.0   Median :0.000   Median :0.000
 Mean   :4.60                       Mean   : 36.0   Mean   :60.1   Mean   :0.302   Mean   :0.581
 3rd Qu.:5.80                       3rd Qu.: 55.5   3rd Qu.:78.5   3rd Qu.:1.000   3rd Qu.:1.000
 Max.   :8.60                       Max.   :100.0   Max.   :89.0   Max.   :1.000   Max.   :3.000
```

A Simple Regression

```
> fit<-with(africa, lm(adrate~muslperc))
> summary(fit)

Call:
lm(formula = adrate ~ muslperc)

Residuals:
    Min      1Q  Median      3Q     Max
-13.828  -5.206   0.279   2.022  23.521

Coefficients:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept)  15.2787     1.8322    8.34 0.00000000023 ***
muslperc     -0.1644     0.0369   -4.45 0.00006390853 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-squared:  0.326, Adjusted R-squared:  0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 0.0000639
```

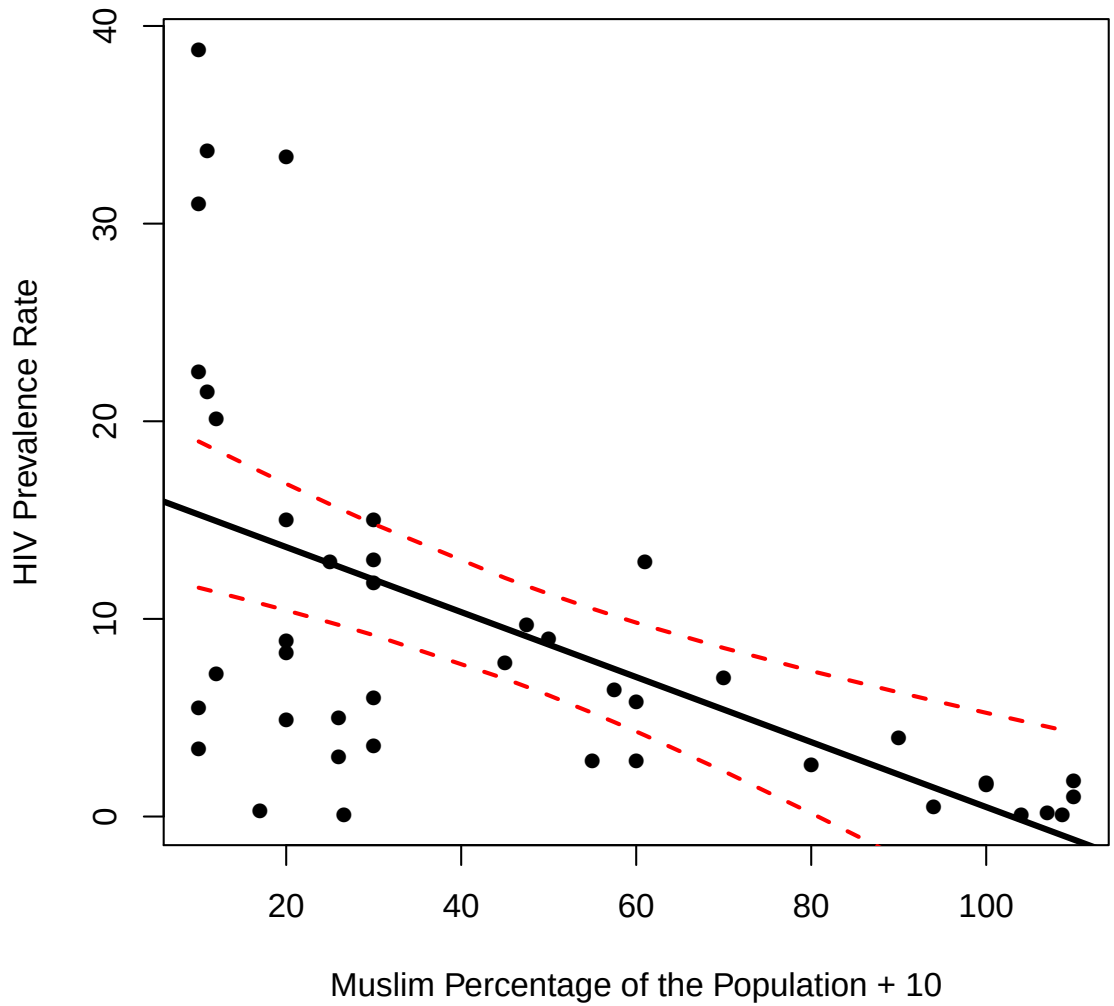
Scatterplot of HIV/AIDS Rates on Muslim Population Percentage, 2001



Adding a Constant to X

```
> africa$muslplusten<-africa$muslperc+10
> fit2<-with(africa, lm(adrate~muslplusten,data=africa))
> summary(fit2)

Call:
lm(formula = adrate ~ muslplusten, data = africa)

Residuals:
    Min      1Q  Median      3Q     Max
-13.828  -5.206   0.279   2.022  23.521

Coefficients:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept)  16.9232     2.1152    8.00 0.00000000066 ***
muslplusten  -0.1644     0.0369   -4.45 0.00006390853 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-squared:  0.326, Adjusted R-squared:  0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 0.0000639
```

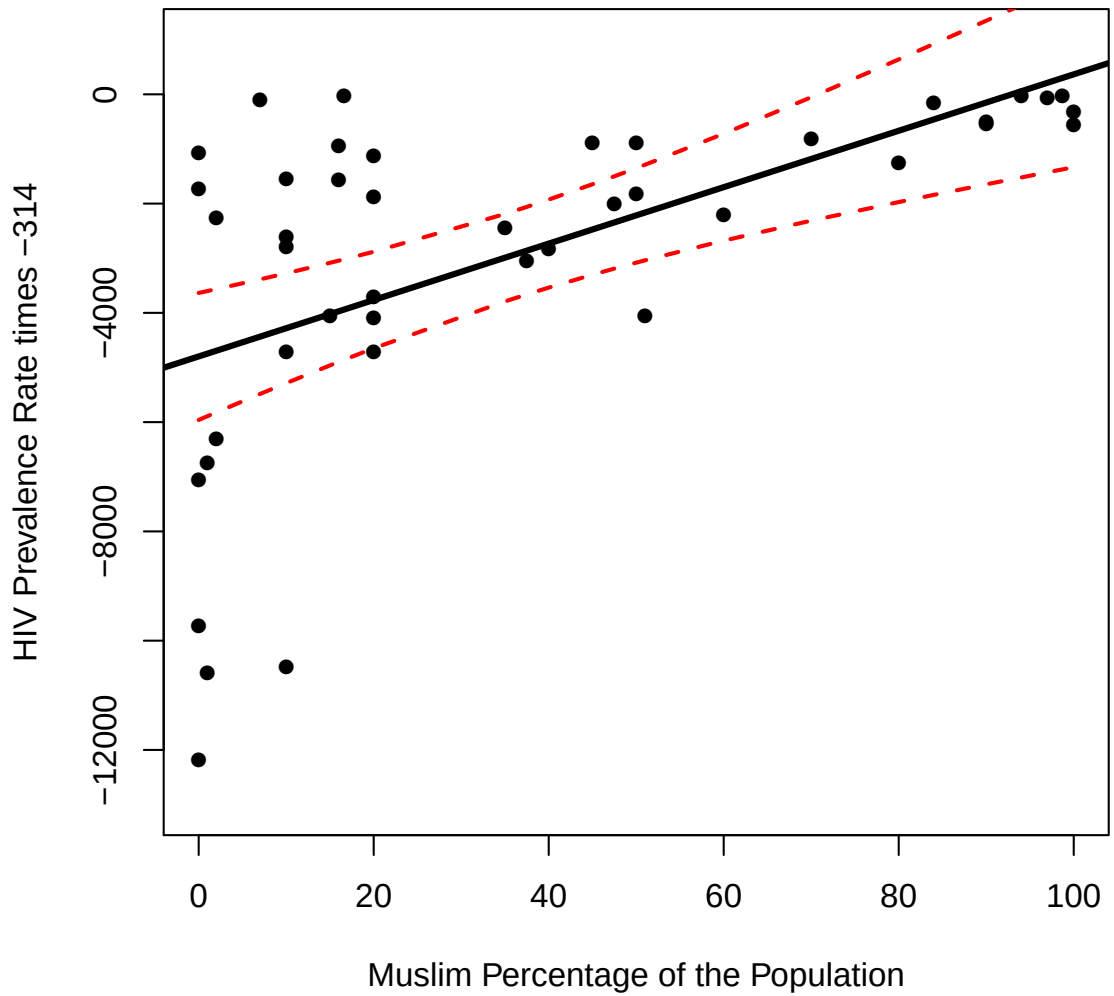
Scatterplot of HIV/AIDS Rates on Rescaled Muslim Population Percentage



Multiplying Y by a Constant

```
> africa$screwyrate<-africa$adrate*(-314)
> fit3<-with(africa, lm(screwyrate~muslperc))
> summary(fit3)

Call:
lm(formula = screwyrate ~ muslperc)

Residuals:
   Min     1Q Median     3Q    Max
 -7386   -635    -88   1635   4342

Coefficients:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept)  -4797.5      575.3   -8.34 0.00000000023 ***
muslperc        51.6       11.6    4.45 0.00006390853 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2600 on 41 degrees of freedom
Multiple R-squared:  0.326, Adjusted R-squared:  0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 0.0000639
```

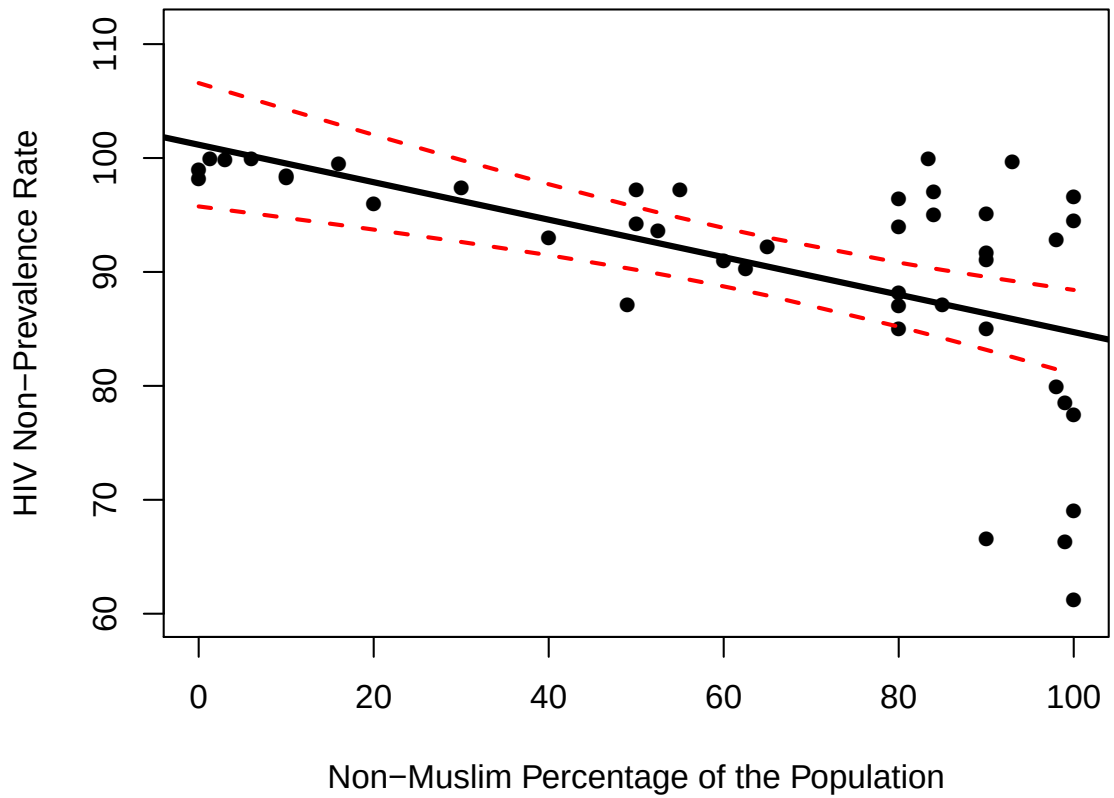
Scatterplot of Rescaled HIV/AIDS Rates on Muslim Population Percentage



Reversing the scales of X and Y

```
> africa$nonmuslimpct <- 100 - africa$muslperc
> africa$noninfected <- 100 - africa$adrate
> fit4<-lm(noninfected~nonmuslimpct,data=africa)
> summary(fit4)

Call:
lm(formula = noninfected ~ nonmuslimpct, data = africa)

Residuals:
    Min      1Q  Median      3Q     Max
-23.521  -2.022  -0.279   5.206  13.828

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  101.1660     2.6808   37.74  < 2e-16 ***
nonmuslimpct  -0.1644     0.0369   -4.45 0.000064 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-squared:  0.326, Adjusted R-squared:  0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 0.0000639
```

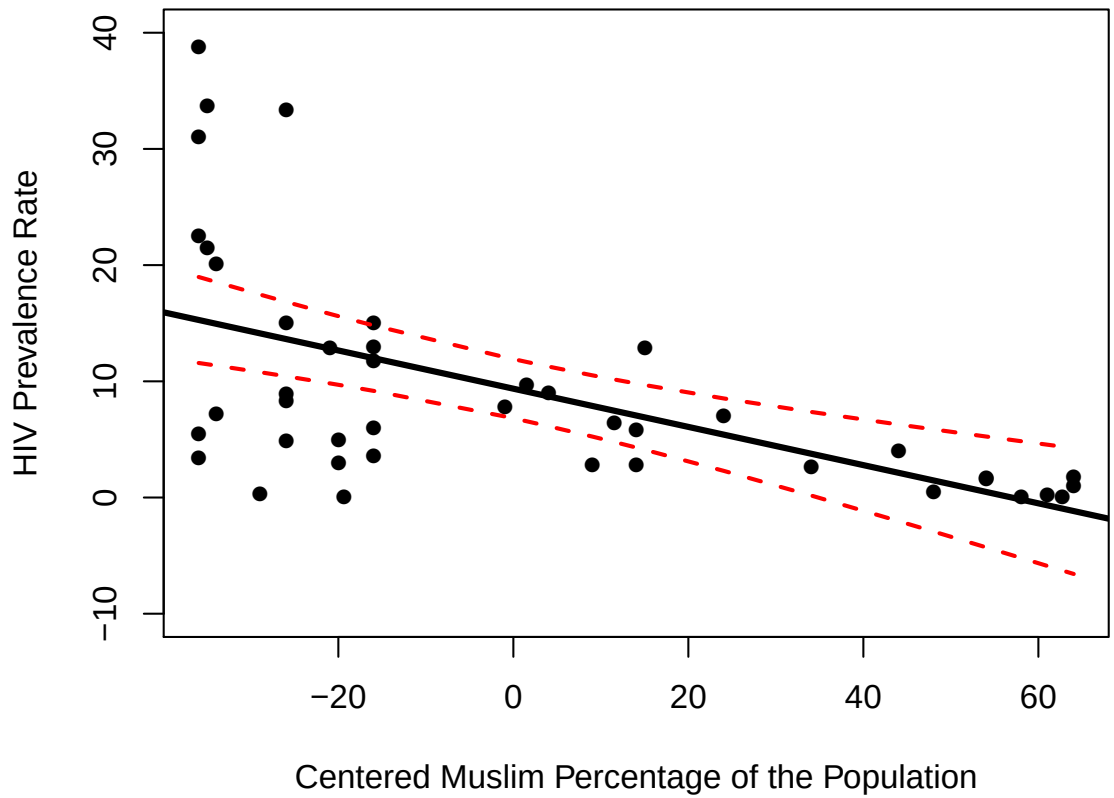
Scatterplot of HIV/AIDS Non-Infection Rates on Non-Muslim Population Percentage



Centering X

```
> africa$muslcenter<-africa$muslperc - mean(africa$muslperc, na.rm=TRUE)
> fit5<-lm(adrate~muslcenter,data=africa)
> summary(fit5)

Call:
lm(formula = adrate ~ muslcenter, data = africa)

Residuals:
    Min      1Q  Median      3Q     Max
-13.828  -5.206   0.279   2.022  23.521

Coefficients:
            Estimate Std. Error t value     Pr(>|t|)
(Intercept)   9.3651     1.2622    7.42 0.0000000042 ***
muslcenter   -0.1644     0.0369   -4.45 0.0000639085 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-squared:  0.326, Adjusted R-squared:  0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 0.0000639
```

Scatterplot of HIV/AIDS Infection Rates on (Centered) Muslim Population Percentage



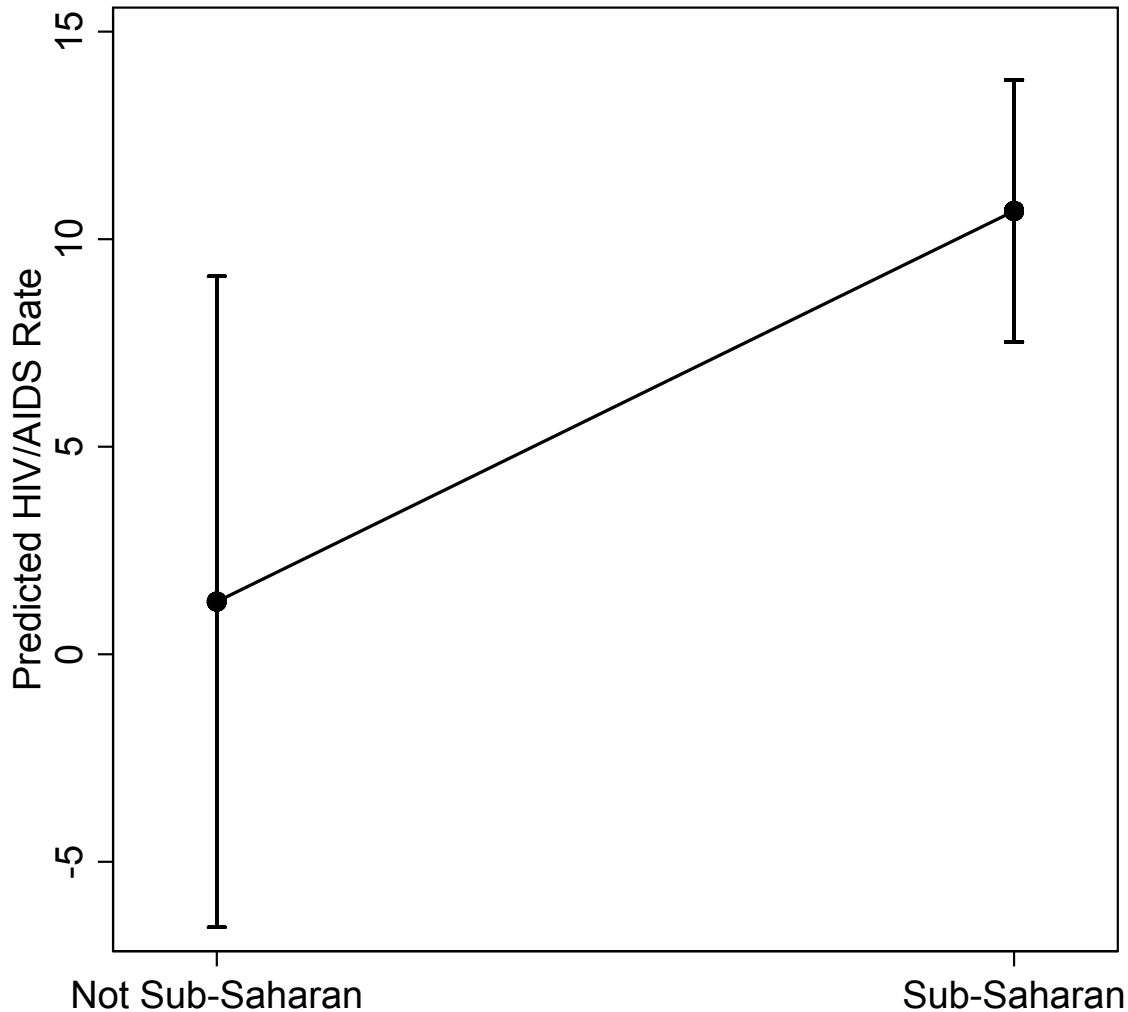
Rescaling X for Interpretability

```
> fit6<-lm(adrate~population,data=africa)
> summary(fit6)

                Estimate    Std. Error t value Pr(>|t|)
(Intercept) 10.5883163475  1.9140361989    5.53 0.000002 ***
population  -0.0000000703  0.0000000671   -1.05      0.3
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.95 on 41 degrees of freedom
Multiple R-squared:  0.0261, Adjusted R-squared:  0.00234
F-statistic:  1.1 on 1 and 41 DF,  p-value: 0.301


> africa$popmil<-africa$population / 1000000
> fit7<-lm(adrate~popmil,data=africa)
> summary(fit7)

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.5883     1.9140    5.53 0.000002 ***
popmil       -0.0703     0.0671   -1.05      0.3
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.95 on 41 degrees of freedom
Multiple R-squared:  0.0261, Adjusted R-squared:  0.00234
F-statistic:  1.1 on 1 and 41 DF,  p-value: 0.301
```

Dichotomous Xs: Bivariate Regression $\equiv$ t-test

```
> fit8<-lm(adrate~subsaharan,data=africa)
> summary(fit8)

Residuals:
   Min     1Q Median     3Q    Max
-10.58  -6.23  -1.78   2.22  28.12

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.27       3.88    0.33     0.75
subsaharanSub-Saharan    9.41       4.19    2.25     0.03 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.51 on 41 degrees of freedom
Multiple R-squared:  0.11, Adjusted R-squared:  0.088
F-statistic: 5.05 on 1 and 41 DF,  p-value: 0.03


> with(africa,
+      t.test(adrate~subsaharan, var.equal=TRUE))

Two Sample t-test

data:  adrate by subsaharan
t = -2.2, df = 41, p-value = 0.03
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -17.8659  -0.9576
sample estimates:
mean in group Not Sub-Saharan     mean in group Sub-Saharan
                        1.267                        10.678
```

Expected Values of HIV/AIDS Infection Rates in Saharan and Sub-Saharan Africa



The results:

```
> fit<-lm(adrate~muslperc, data=africa)
> summary.lm(fit)

Call:
lm(formula = adrate ~ muslperc, data = africa)

Residuals:
    Min      1Q  Median      3Q     Max
-13.828  -5.206   0.279   2.022  23.521

Coefficients:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept)  15.2787     1.8322    8.34 0.00000000023 ***
muslperc     -0.1644     0.0369   -4.45 0.00006390853 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-squared:  0.326, Adjusted R-squared:  0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 0.0000639
```

Reporting

The table:

Table: OLS Regression Model of HIV/AIDS Rates in Africa, 2001

Variables	Model I
(Constant)	15.28
	(1.83)
Muslim Percentage of the Population	-0.164*
	(0.037)
Adjusted R^2	0.31

Note: $N = 43$. Cell entries are coefficient estimates; numbers in parentheses are estimated standard errors. Asterisks indicate $p < .05$ (one-tailed). See text for details.

Some Guidelines ("Rules"?)

Tables:

- *Use column headings descriptively.*
- *Use multiple rows / columns rather than multiple tables.*
- *Learn about significant digits, and don't report more than 4-5 of them.*
- *Use a figure to replace a table when you can.*
- *Be aware of norms about *s.*

Figures:

- *Report the scale of axes, and label them.*
- *Use as much "space" as you need, but no more.*
- *Use color sparingly.*