

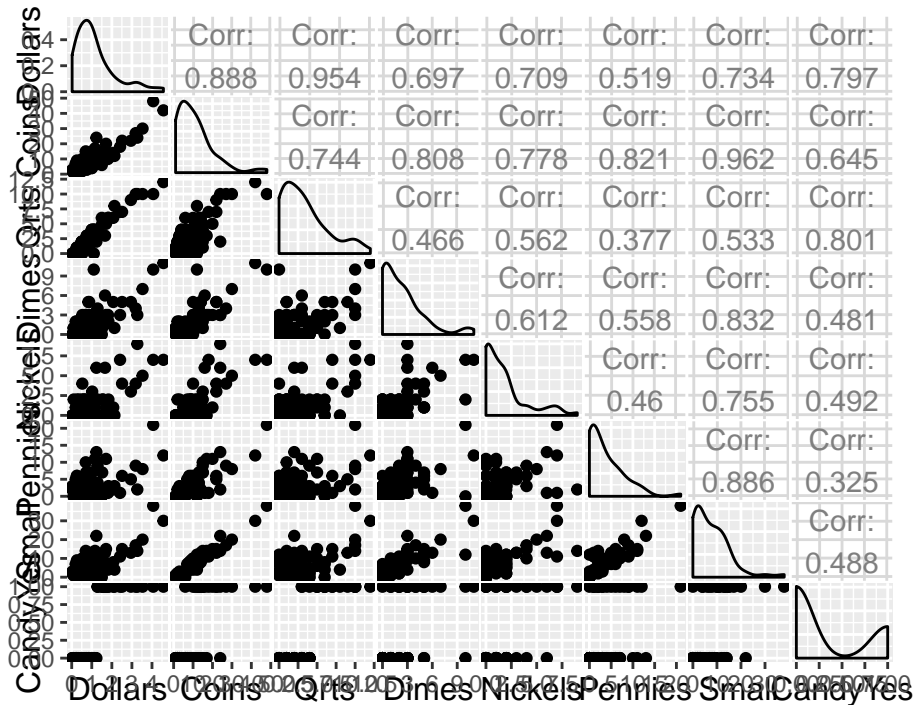
### Multicollinearity

Consider the following data set collected from church offering plates in 62 consecutive Sundays. Also noted is whether there was enough change to buy a candy bar for \$1.25.

```
library(readr)
library(ggplot2)
library(GGally)
library(dplyr)
library(xtable)
Offering <- read_csv("~/Dropbox/teaching/coins.csv")

Offering <- Offering %>%
  mutate(Small= Dimes + Nickels + Pennies) %>%
  mutate(CandyYes = ifelse(Dollars > 1.25, 1, 0))

ggpairs(Offering)
```



## Sequential models

```
summary(glm(CandyYes ~ Coins, data = Offering, family="binomial"))$coef
##           Estimate Std. Error  z value    Pr(>|z|)
## (Intercept) -4.1397779  0.9956669 -4.157794 3.213356e-05
## Coins       0.2857377  0.0771673  3.702834 2.132040e-04

summary(glm(CandyYes ~ Small, data = Offering, family="binomial"))$coef
##           Estimate Std. Error  z value    Pr(>|z|)
## (Intercept) -2.3272968  0.58492766 -3.978777 6.927062e-05
## Small       0.1836372  0.05756784  3.189927 1.423088e-03

summary(glm(CandyYes ~ Coins + Small, data = Offering, family="binomial"))$coef
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
##           Estimate Std. Error  z value    Pr(>|z|)
## (Intercept) -16.975017  7.801826 -2.175775 0.02957210
## Coins       3.489612  1.749639  1.994476 0.04610008
## Small      -3.042380  1.572702 -1.934493 0.05305256
```

Notice that the directionality of the low coins changes when it is included in the model that already contains the number of coins total. Moral of the story: be very very very careful interpreting coefficients when you have multiple explanatory variables.