Missing Data Analysis

William Murrah

Data we will use:

```
#  ?nhanes
data("nhanes2")
```

```
Warning in data("nhanes2"): data set 'nhanes2' not found
```

```
achieve <- read.csv('data/Achieve.csv')
achieve <- achieve[ ,c("geread", "gevocab", "gender", "age")]
achieve$gender <- achieve$gender - 1
achieve$age <- achieve$age/12
```

```
mod <- lm(geread ~ gevocab + gender + age, data = achieve)
summary(mod)
```

```
Call:
lm(formula = geread ~ gevocab + gender + age, data = achieve)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3433 -1.1259 -0.4311  0.6090  8.6165

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.195695   0.416563   7.672 1.85e-14 ***
gevocab      0.528256   0.008191  64.489  < 2e-16 ***
gender       0.038136   0.038829   0.982  0.32605
age         -0.139266   0.046113  -3.020  0.00253 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.966 on 10316 degrees of freedom
Multiple R-squared:  0.2898,    Adjusted R-squared:  0.2896
F-statistic:  1403 on 3 and 10316 DF,  p-value: < 2.2e-16
```

```
# b0 <-  1.97
# b1 <-  0.53
# b2 <-  0.04
# n <- 10320
# sigma <- 1.97
#
# set.seed(123)
# gevocab <- rnorm(n = n, mean = 4.94, sd = 2.37)
# gender <- rep(0:1, n/2)
#
```

```
# geread <- b0 + b1*gevocab + b2*gender + rnorm(n = n, mean = 0, sd = sigma )
# simdata <- data.frame(read = geread,
#                       vocab = gevocab,
#                       female = factor(gender,
#                                       labels = c("male", "female")))
```

```
# logistic <- function(x) exp(x)/(1 + exp(x))
# inv.logit()
# set.seed(1234)
# r.mcar <- 1 - rbinom(n, 1, 0.50)
# r.mar <- 1 - rbinom(n, 1, logistic(simdata$vocab))
# set.seed(1234)
# r.mnar <- 1 - rbinom(n, 1, inv.logit(simdata$read))
```

```
R version 3.2.2 (2015-08-14)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.3 LTS

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] knitr_1.11       mosaic_0.11      mosaicData_0.9.1 car_2.1-0
[5] ggplot2_1.0.1    lattice_0.20-33  dplyr_0.4.3

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.1         formatR_1.2.1       nloptr_1.0.4
 [4] plyr_1.8.3          tools_3.2.2         digest_0.6.8
 [7] lme4_1.1-10         evaluate_0.8        gtable_0.1.2
[10] nlme_3.1-122        mgcv_1.8-7          Matrix_1.2-2
[13] DBI_0.3.1           yaml_2.1.13         parallel_3.2.2
[16] SparseM_1.7         ggdendro_0.1-17     proto_0.3-10
[19] gridExtra_2.0.0     stringr_1.0.0       MatrixModels_0.4-1
[22] grid_3.2.2          nnet_7.3-11         R6_2.1.1
[25] rmarkdown_0.8.1     minqa_1.2.4         reshape2_1.4.1
[28] magrittr_1.5        scales_0.3.0        htmltools_0.2.6
[31] MASS_7.3-44         splines_3.2.2       assertthat_0.1
[34] pbkrtest_0.4-2      colorspace_1.2-6    quantreg_5.19
[37] stringi_0.5-5       munsell_0.4.2
```