Lab 2

Tziporah Horowitz

11:59PM February 17, 2020

More Basic R Skills

- Calculate the average of 1000 realizations of Bernoullis with p = 0.9 in one line using `rbinom`.

```
mean(rbinom(1000, 1, .9))
```

```
## [1] 0.913
```

- In class we considered a variable `x_3` which measured "criminality". We imagined L = 4 levels "none", "infraction", "misdimeanor" and "felony". Create a variable `x3` here with 100 random elements (equally probable). Create it as a nominal (i.e. unordered) factor.

```
x_3 <- factor(sample(c("none", "infraction", "misdimeanor", "felony"), 100, replace = TRUE))
```

- Convert this variable into three binary variables without any information loss and put them into a data matrix.

```
X <- matrix(data = NA, length(x_3), 3)
X[ , 1] <- as.numeric(x_3 == "infraction")
X[ , 2] <- as.numeric(x_3 == "misdimeanor")
X[ , 3] <- as.numeric(x_3 == "felony")
colnames(X) <- c("is_infraction", "is_misdimeanor", "is_felony")
head(X)
```

```
##      is_infraction is_misdimeanor is_felony
## [1,]             0              0         1
## [2,]             0              0         0
## [3,]             0              0         0
## [4,]             0              1         0
## [5,]             1              0         0
## [6,]             1              0         0
```

- What should the sum of each row be (in English)? Verify that.

It should be 1 or 0 because the categories are mutually exclusive.

```
rowSums(X)
```

```
##   [1] 1 0 0 1 1 1 1 0 0 1 1 1 1 1 1 0 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 0
##  [36] 0 1 0 1 1 0 1 0 0 1 1 0 1 1 1 0 1 1 0 0 1 0 1 1 1 1 0 1 0 1 1 1 0 1 1
##  [71] 1 1 1 1 0 1 1 1 1 1 0 0 1 1 1 1 1 1 0 1 1 0 1 0 0 1 0 1 1 0
```

- How should the column sum look (in English)? Verify that.

```r
colSums(X)
```

```
##  is_infraction is_misdimeanor      is_felony
##             20             25             23
```

- Generate a matrix with 100 rows where the first column is realization from a normal with mean 17 and variance 38, the second column is uniform between -10 and 10, the third column is poisson with mean 6, the fourth column in exponential with lambda of 9, the fifth column is binomial with n = 20 and p = 0.12 and the sixth column is a binary variable with exactly 24% 1's dispersed randomly. Name the columns based on the r.v. Name the rows the entries of the `fake_first_names` vector.

```r
fake_first_names = c(
  "Sophia", "Emma", "Olivia", "Ava", "Mia", "Isabella", "Riley",
  "Aria", "Zoe", "Charlotte", "Lily", "Layla", "Amelia", "Emily",
  "Madelyn", "Aubrey", "Adalyn", "Madison", "Chloe", "Harper",
  "Abigail", "Aaliyah", "Avery", "Evelyn", "Kaylee", "Ella", "Ellie",
  "Scarlett", "Arianna", "Hailey", "Nora", "Addison", "Brooklyn",
  "Hannah", "Mila", "Leah", "Elizabeth", "Sarah", "Eliana", "Mackenzie",
  "Peyton", "Maria", "Grace", "Adeline", "Elena", "Anna", "Victoria",
  "Camilla", "Lillian", "Natalie", "Jackson", "Aiden", "Lucas",
  "Liam", "Noah", "Ethan", "Mason", "Caden", "Oliver", "Elijah",
  "Grayson", "Jacob", "Michael", "Benjamin", "Carter", "James",
  "Jayden", "Logan", "Alexander", "Caleb", "Ryan", "Luke", "Daniel",
  "Jack", "William", "Owen", "Gabriel", "Matthew", "Connor", "Jayce",
  "Isaac", "Sebastian", "Henry", "Muhammad", "Cameron", "Wyatt",
  "Dylan", "Nathan", "Nicholas", "Julian", "Eli", "Levi", "Isaiah",
  "Landon", "David", "Christian", "Andrew", "Brayden", "John",
  "Lincoln"
)

n <- 100
X <- matrix(data = NA, n, 6)
X[ , 1] <- rnorm(n, 17, sqrt(38))
X[ , 2] <- runif(n, -10, 10)
X[ , 3] <- rpois(n, 6)
X[ , 4] <- rexp(n, 9)
X[ , 5] <- rbinom(n, 20, .12)
X[ , 6] <- sample(c(rep(1, n*.24), rep(0, n*.76)))

rownames(X) <- fake_first_names
head(X)
```

```
##              [,1]      [,2] [,3]       [,4] [,5] [,6]
## Sophia   21.01231 -1.502374    7 0.06881794    0    0
## Emma     17.85397  5.169649    8 0.03413744    3    1
## Olivia   18.23009 -1.783604    4 0.11863330    2    0
## Ava      16.97467  6.653965    2 0.13590157    1    0
## Mia      21.13994  3.074378    7 0.31915589    0    0
## Isabella 15.76385 -4.842958    6 0.00188204    3    0
```

- Create a data frame of the same data as above except make the binary variable a factor "DOMESTIC" vs "FOREIGN" for 0 and 1 respectively. Print out the top few rows to check this worked correctly.

```
df <- data.frame(X)
df$X6 <- factor(df$X6, levels = c(0, 1), labels = c("DOMESTIC", "FOREIGN"))
head(df)
```

```
##                X1        X2 X3         X4 X5       X6
## Sophia   21.01231 -1.502374  7 0.06881794  0 DOMESTIC
## Emma     17.85397  5.169649  8 0.03413744  3  FOREIGN
## Olivia   18.23009 -1.783604  4 0.11863330  2 DOMESTIC
## Ava      16.97467  6.653965  2 0.13590157  1 DOMESTIC
## Mia      21.13994  3.074378  7 0.31915589  0 DOMESTIC
## Isabella 15.76385 -4.842958  6 0.00188204  3 DOMESTIC
```

- Print out a table of the binary variable. Then print out the proportions of "DOMESTIC" vs "FOREIGN".

```
table(df$X6)/n
```

```
##
## DOMESTIC  FOREIGN
##     0.76     0.24
```

- Print out a summary of the whole dataframe.

```
summary(df)
```

```
##        X1                X2                X3              X4
##  Min.   :-0.8024   Min.   :-9.96134   Min.   : 1.00   Min.   :0.0005653
##  1st Qu.:13.5986   1st Qu.:-4.63860   1st Qu.: 4.00   1st Qu.:0.0277057
##  Median :17.3896   Median :-0.39104   Median : 6.00   Median :0.0754768
##  Mean   :17.3459   Mean   : 0.03417   Mean   : 6.02   Mean   :0.1057165
##  3rd Qu.:21.6023   3rd Qu.: 4.97025   3rd Qu.: 7.00   3rd Qu.:0.1471859
##  Max.   :31.7354   Max.   : 9.71040   Max.   :13.00   Max.   :0.4230400
##        X5              X6
##  Min.   :0.00   DOMESTIC:76
##  1st Qu.:1.00   FOREIGN :24
##  Median :2.00
##  Mean   :2.47
##  3rd Qu.:3.00
##  Max.   :8.00
```

- Let `n = 50`. Create a n x n matrix `R` of exactly 50% entries 0's, 25% 1's 25% 2's. These values should be in random locations.

```
n = 50
R <- matrix(sample(c(rep(0, n^2 * .5), c(rep(1, n^2 * .25)), c(rep(2, n^2 * .25)))), n, n)
table(R)
```

```
## R
##    0    1    2
## 1250  625  625
```

- Randomly punch holes (i.e. `NA`) values in this matrix so that each entry is missing with probability 30%.

```
for (i in 1 : n){
  for (j in 1 : n){
    if (runif(1) < 0.3){
      R[i, j] <- NA
    }
  }
}
head(R)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## [1,]   NA   NA    0    0   NA    2    0    2   NA     0     1     1     2
## [2,]    1    0    0    0    2   NA    0    0    2     0    NA     2     1
## [3,]   NA    0    2   NA    0    0    1    1    0    NA    NA     0     1
## [4,]   NA    2    0    0   NA    2   NA    0   NA     0     0    NA     2
## [5,]    0    0   NA    0    0    1    2    2    0     1     2     0    NA
## [6,]    1    2    0    0    2    1   NA    0    0     0     0    NA     2
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
## [1,]     2     0    NA     2    NA     0     0     0     1     1     0
## [2,]     0    NA     2     1     0     2    NA     2     2     2     1
## [3,]    NA    NA    NA     2     2     1     2     1     1    NA     1
## [4,]    NA     1     2     1    NA    NA     0    NA     0    NA     1
## [5,]     1     0     1     0    NA    NA    NA    NA    NA    NA     2
## [6,]    NA     1     0    NA     0     0     0     0    NA     0    NA
##      [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35]
## [1,]     0     2     1    NA     2    NA    NA    NA    NA     0     0
## [2,]     0     1    NA    NA     0     0     0     0     0     0     2
## [3,]     0     1    NA     2     2    NA     2    NA     2     2     0
## [4,]     0     0     0     0    NA     2     2    NA    NA     0    NA
## [5,]     2     0     2     0     0     0    NA    NA     1     0     1
## [6,]     2     0     0     2     2     1    NA    NA     0     2    NA
##      [,36] [,37] [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46]
## [1,]    NA    NA     0     0     2     0     1     0     1     0     2
## [2,]     0     2     2    NA     0     2    NA    NA     0     2    NA
## [3,]     0    NA     0    NA    NA     1     0     0     0     0     0
## [4,]     1     0     0     0     0    NA    NA    NA     0     1    NA
## [5,]     2     0    NA     0     0     0    NA     1    NA    NA     2
## [6,]     2    NA     0     0    NA     2     1     0     2     1     0
##      [,47] [,48] [,49] [,50]
## [1,]     2     2     2     2
## [2,]     0     0     0    NA
## [3,]     0     0     1    NA
## [4,]    NA    NA     0    NA
## [5,]     1    NA     0     2
## [6,]     0     2     2     0
```

```
sum(is.na(R)) / n^2
```

```
## [1] 0.296
```

- Sort the rows in matrix `R` by the largest row sum to lowest. Be careful about the NA's!

```r
R[order(rowSums(R, na.rm = TRUE), decreasing = TRUE), ]
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
##  [1,]   NA   NA   NA    0    0    1    2    0    0     0     2     0     1
##  [2,]    0    2    1    0    2    0    2    0   NA     2     2     0     0
##  [3,]    2    2    2   NA   NA    2   NA   NA    0    NA     2     1    NA
##  [4,]    2    2   NA    0   NA    2    0    2   NA    NA     1     0     0
##  [5,]    2    0    0    2    0   NA    0    1    0     0    NA     1     2
##  [6,]    0    1   NA    0    2    1    1    0    2     2    NA     2    NA
##  [7,]    1    2    0    0    0    2   NA    2   NA     0    NA    NA    NA
##  [8,]   NA   NA    0    0   NA    2    0    2   NA     0     1     1     2
##  [9,]    0    2    0    1    1    2    0    0   NA    NA     2     1     0
## [10,]    2    1    2    0    0    1   NA    1    1     0    NA     0     0
## [11,]    1    0    0    0    2   NA    0    0    2     0    NA     2     1
## [12,]    1    1    0    2   NA    0    2    2   NA    NA    NA    NA     1
## [13,]   NA    1    2    1    1   NA    0    2   NA     0    NA     2    NA
## [14,]    1    2    0    0    2    1   NA    0    0     0     0    NA     2
## [15,]   NA    2    0    0    1    2    2    1    0     1    NA     1     2
## [16,]   NA    2   NA    2   NA    0    0   NA    0     2     0     0     0
## [17,]    1    0    0    1   NA    0   NA    0    0     2     0     2     2
## [18,]   NA    0    2   NA    0    0    1    1    0    NA    NA     0     1
## [19,]    2   NA    1    0    1    2    2    0    1     2     0     1     0
## [20,]    0    1    0    2   NA    2   NA    2    2     2     0     0    NA
## [21,]   NA    0   NA    0    1    1    2    0    0     0     1     1     1
## [22,]    0    0    0   NA    0   NA    1   NA    2     0     2     2     1
## [23,]   NA   NA   NA    0    2    2    1    2    2     0    NA    NA     1
## [24,]   NA    0    0   NA    0    2   NA    1    0     0     1     0    NA
## [25,]    0    0   NA    0    0    1    2    2    0     1     2     0    NA
## [26,]   NA    1    0    2   NA   NA   NA    1    2     0     0     0     0
## [27,]    0    1    1    1    2    1    2    0    1    NA    NA     0     0
## [28,]    1    2   NA    0   NA   NA   NA    1   NA     0     0     1    NA
## [29,]   NA    0   NA    1    1   NA    0    0    1    NA     1    NA     0
## [30,]    0    0   NA    1    1    0    0   NA   NA     2     0     1     0
## [31,]    2    2    0    0    0    1   NA    2   NA     2    NA     2    NA
## [32,]    0    1    0   NA   NA   NA    2    0   NA     0    NA     2     1
## [33,]    1    2    0   NA    0    0    1    0    1    NA     2    NA     1
## [34,]    0    1    0    0    1   NA   NA   NA    2     1     0     0    NA
## [35,]    0   NA    1   NA    0    0    1   NA    1     1     2     2    NA
## [36,]    2    1    1   NA    0    0    2    0    0     2     1    NA     1
## [37,]   NA    2   NA   NA    1    0   NA   NA   NA    NA     0     0    NA
## [38,]    0    1   NA   NA    2    2   NA   NA    0     2     0    NA     0
## [39,]    0    0   NA    0    0    0    1    1   NA     0     1     2     0
## [40,]   NA    1    2    1    2    1    0    0    1     0     2     2     1
## [41,]   NA    0   NA   NA    1   NA    1    1    1     0     1     1     0
## [42,]   NA    0    2    0    0   NA    0    0   NA     1    NA     0     2
## [43,]    0    2   NA    0    2   NA    2    2    0     1    NA     2     0
## [44,]    0   NA    0    2    0    0    1   NA    0    NA    NA     0     0
## [45,]   NA    2    1    2    1   NA    0    0   NA    NA    NA     0    NA
## [46,]   NA    0    0    2   NA    0    2    0   NA    NA     2     1     1
## [47,]    1   NA    0    0    0    0   NA    1   NA     0     0     0     1
## [48,]   NA    2    0    0   NA    2   NA    0   NA     0     0    NA     2
## [49,]    0   NA    1    0   NA   NA    0    2    0     1    NA     0    NA
## [50,]   NA    0    0    0    2   NA   NA    0   NA     1    NA     0     1
```

```
##       [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
##  [1,]     1    NA     1     0     2    NA     0    NA     1     2     1
##  [2,]     1     2     1     1     0     1    NA    NA     0     2     2
##  [3,]     0    NA     0     1     0     2     1    NA     0    NA     0
##  [4,]     2     0     1     1     1     0     0     0     0     0    NA
##  [5,]     2     2     1     1    NA     2    NA     0    NA    NA    NA
##  [6,]     1     1     0     1    NA    NA     1     1    NA     1    NA
##  [7,]    NA     0     2     2     0     0     2     2     2    NA     1
##  [8,]     2     0    NA     2    NA     0     0     0     1     1     0
##  [9,]    NA     0    NA    NA     1    NA    NA     2     1    NA     0
## [10,]     0    NA     0     2     0     0     1     2     0     2     0
## [11,]     0    NA     2     1     0     2    NA     2     2     2     1
## [12,]    NA     2     0    NA     0    NA     1     0     0     2     0
## [13,]     0     0     1     0     0     0     0    NA    NA    NA    NA
## [14,]    NA     1     0    NA     0     0     0     0    NA     0    NA
## [15,]    NA     1    NA     0     0     2     1     2     2     0    NA
## [16,]     2    NA     0     2     2    NA     1    NA     0    NA     2
## [17,]    NA    NA    NA    NA     2    NA     2     1     2     1     0
## [18,]    NA    NA    NA     2     2     1     2     1     1    NA     1
## [19,]    NA     0    NA    NA    NA    NA    NA     0     0    NA     1
## [20,]     0     2     2     0     0     0     2     0     0    NA    NA
## [21,]    NA    NA     0     0     2    NA     0     2    NA    NA     0
## [22,]    NA     0     0     0    NA     2     0    NA     1     2     2
## [23,]     2     0     2     0     0     2     0     1     0    NA    NA
## [24,]     1     1     2     1    NA     2     2     1     1    NA     0
## [25,]     1     0     1     0    NA    NA    NA    NA    NA    NA     2
## [26,]     2     2     1     2     0    NA     1     1     1    NA    NA
## [27,]    NA    NA     0    NA     0    NA     0     1     1    NA    NA
## [28,]     0     0    NA     2     0     2     2     0     0    NA     2
## [29,]    NA    NA     0     1     2     1     0     1     0     2     2
## [30,]    NA    NA    NA    NA     2     0     2     0     0     0     2
## [31,]     0    NA     2    NA    NA     1    NA    NA     2     1     0
## [32,]     0     2     0    NA     0     2     0    NA     0     0     1
## [33,]     1    NA     1     0    NA     2     0     1     0     0     0
## [34,]     0     2     2    NA     0     2     2    NA    NA     2     0
## [35,]     0     0     0    NA     0     0    NA     2    NA     2     0
## [36,]     1    NA    NA    NA     0     0    NA     0     1    NA    NA
## [37,]     0     0     0     0    NA     1     2     1     0     2     0
## [38,]    NA    NA     0     2     0    NA     0    NA     0     0     0
## [39,]     0     2     0     1     0     1     0    NA     2    NA     1
## [40,]     1     0     2     0     1     0     0    NA     0     1     1
## [41,]     1     0     0    NA     0     0     0     0     0     2     0
## [42,]    NA    NA     0    NA     0    NA    NA     0     1    NA     2
## [43,]    NA     0     0    NA     1     0     0     0    NA     0     0
## [44,]    NA     1     0    NA     0     2     1     0    NA     2     0
## [45,]     0     0     0     0     1     1    NA    NA     0    NA     1
## [46,]     0     2     0    NA     0     1     0     0    NA     0     0
## [47,]    NA     1     0     0     0     2     0     0    NA     0    NA
## [48,]    NA     1     2     1    NA    NA     0    NA     0    NA     1
## [49,]     0    NA     0     0     2    NA     2     0     1     0     0
## [50,]     1     0     0     1     2     0     0     0    NA     1     1
##       [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35]
##  [1,]     0    NA     2     1    NA     2     2     0     1     2     1
##  [2,]    NA    NA     0     1    NA     1    NA     0     0     0     2
```

```
##  [3,]    NA     2     0    NA     2     1     2     2     1     0    NA
##  [4,]     2     0    NA     0    NA     2     1     1     0     0     0
##  [5,]    NA    NA     1     0     1     2     0     2    NA    NA     0
##  [6,]    NA     0    NA    NA     1     2    NA    NA     2     0     0
##  [7,]     1    NA     0     2     2     0     0    NA     1    NA    NA
##  [8,]     0     2     1    NA     2    NA    NA    NA    NA     0     0
##  [9,]     2    NA     0    NA     2     2     2     1     0     0    NA
## [10,]     0    NA     2    NA     1     2     0    NA     2     2     0
## [11,]     0     1    NA    NA     0     0     0     0     0     0     2
## [12,]     0     0     2     1     2     1     0     0     1     0     1
## [13,]     0     0     1     1     2     0    NA    NA     2     2     0
## [14,]     2     0     0     2     2     1    NA    NA     0     2    NA
## [15,]     1     0     2    NA     1     0     2    NA     1    NA    NA
## [16,]     0    NA     0     0     0    NA     2    NA     2    NA     1
## [17,]     0    NA     0     0    NA     0    NA     2     2    NA     0
## [18,]     0     1    NA     2     2    NA     2    NA     2     2     0
## [19,]     0    NA     2     2    NA     0     1     1     1     0     1
## [20,]     0     0     1     2     2    NA     0     0     1     0     0
## [21,]     1     0     0     0     2     0     0    NA     2    NA    NA
## [22,]    NA    NA     2    NA     0     0     2     2    NA    NA    NA
## [23,]     0    NA     0     0     0     2     2    NA     0     1     0
## [24,]     0    NA    NA    NA     1     0    NA    NA    NA     2     1
## [25,]     2     0     2     0     0     0    NA    NA     1     0     1
## [26,]    NA     2    NA    NA     1     0    NA    NA     0     2    NA
## [27,]     2     1    NA    NA    NA     0    NA     1     1     0     1
## [28,]    NA     1    NA    NA     0    NA     0     0    NA    NA    NA
## [29,]     1    NA     1     2     0     0     2     0     1     1    NA
## [30,]    NA     1     1     2    NA     1     1    NA     0     0     0
## [31,]     0    NA     0     1    NA    NA     1     0     0    NA     0
## [32,]     0     2     0    NA    NA     2     0    NA     1     1     2
## [33,]    NA     0     0     0    NA    NA     2     1    NA     0     0
## [34,]     1     0     1     2     0    NA     0     2    NA     1     0
## [35,]    NA     0    NA    NA     0    NA    NA    NA     0     0    NA
## [36,]    NA     0     0     0     0     1     0     0    NA     1    NA
## [37,]    NA     0    NA    NA    NA    NA     0     2    NA    NA     1
## [38,]    NA     0    NA     0     1    NA    NA     1    NA    NA     0
## [39,]     0     0    NA     0     1     0     2     0    NA     1     1
## [40,]     0     0     1     0    NA    NA    NA     0    NA     0    NA
## [41,]    NA     0     2     1    NA     0     0     1     0    NA     2
## [42,]    NA     0     0    NA    NA     2     1     0     1    NA    NA
## [43,]    NA     0     2    NA     0    NA     2     0    NA     2     0
## [44,]     0    NA     0     0     2    NA    NA     0    NA     0     0
## [45,]     1     2    NA     0    NA     1     0     0     0     0    NA
## [46,]    NA     0     2    NA     1    NA     1     0    NA     0     1
## [47,]     1     1     1    NA     2    NA     0     0     2     2     0
## [48,]     0     0     0     0    NA     2     2    NA    NA     0    NA
## [49,]     0     0     1    NA    NA     0     0    NA     0     0    NA
## [50,]    NA     1    NA    NA     1    NA     0     0    NA     0    NA
##        [,36] [,37] [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46]
##  [1,]    NA     1     0    NA     0     1     2    NA    NA     2     1
##  [2,]     0     2     1     0     0     1     1    NA     1     1     2
##  [3,]     2    NA    NA     1     0     2     0     0     1     1     1
##  [4,]     0     1     2     2     2     1     1     2    NA     2    NA
##  [5,]     2     1     1    NA     1     1    NA     2     0     2     0
```

```
##  [6,]     1     1     2     2    NA     0     0     2     2     0     1
##  [7,]    NA    NA     0    NA     2    NA     0     1     0     1     2
##  [8,]    NA    NA     0     0     2     0     1     0     1     0     2
##  [9,]     2    NA     0     1    NA     2     1    NA     1     2    NA
## [10,]     2    NA    NA     2    NA     0     0     0     0     0     2
## [11,]     0     2     2    NA     0     2    NA    NA     0     2    NA
## [12,]     2     1     0     1     0     2     0     2    NA     0    NA
## [13,]     0     2     1     0     2     0     2     2     2    NA     0
## [14,]     2    NA     0     0    NA     2     1     0     2     1     0
## [15,]    NA     0    NA     1     0    NA     0    NA     0     0     0
## [16,]     0     2     1    NA     2     0    NA     2     2     1     0
## [17,]     0     0     1     1    NA     2    NA     1     2    NA    NA
## [18,]     0    NA     0    NA    NA     1     0     0     0     0     0
## [19,]     2     0     0     0     2     1    NA     0    NA     1    NA
## [20,]    NA    NA    NA    NA    NA     1     1     0    NA     2    NA
## [21,]    NA     2     2     1     1    NA     2     2    NA    NA     1
## [22,]     0     0     0     1    NA     0     0     0     1     1     2
## [23,]     2     1     2     0    NA    NA     0    NA     0     0    NA
## [24,]     1     0     0     0    NA     1    NA     2     2    NA    NA
## [25,]     2     0    NA     0     0     0    NA     1    NA    NA     2
## [26,]     2     0     0    NA    NA     0     1    NA     0     1     0
## [27,]     1     0     2    NA     0    NA     0     0     0     2     0
## [28,]     1     1     0     0     2    NA     0     2     1     0     1
## [29,]     0    NA     1    NA     1     0     0     0    NA    NA    NA
## [30,]    NA     1     0     0     0    NA     1     1    NA     1     1
## [31,]    NA     0    NA     0    NA     1     0     1     0    NA     0
## [32,]    NA     1     2     0     0     0     0    NA    NA     2    NA
## [33,]    NA    NA    NA     2     0     0     2     2    NA    NA    NA
## [34,]     0     0     0     0    NA     1     0    NA     1     1    NA
## [35,]     0    NA     2     0     2     1    NA     1     2    NA     0
## [36,]    NA     0    NA    NA    NA     1     0     0     2     1     0
## [37,]     2    NA     2     2     1     1     0     1     0     0     2
## [38,]     1     0    NA    NA     2     0     2     2    NA     2     1
## [39,]     1     1     1    NA     0     0     1     0    NA     1     1
## [40,]     0     0     0     0    NA     0    NA     1     0     0     0
## [41,]     1    NA     0    NA     1     0    NA     2    NA    NA    NA
## [42,]    NA     1     1    NA     1     0    NA     1    NA     0     2
## [43,]     0    NA    NA     0    NA     0     0     0     0    NA     0
## [44,]     0     0     2     0     0     1     1     1     2     0     0
## [45,]     0     1     1    NA     0     0     1    NA    NA     1     2
## [46,]     0    NA     1    NA     2    NA     0    NA     0    NA    NA
## [47,]     1     0    NA    NA     0     0     0    NA     2     0     0
## [48,]     1     0     0     0     0    NA    NA    NA     0     1    NA
## [49,]    NA     0     1     1    NA     0     1     2     0    NA     0
## [50,]    NA     0    NA     0     0    NA    NA     0     1    NA     0
##       [,47] [,48] [,49] [,50]
##  [1,]     0     2     1    NA
##  [2,]     0     1    NA     0
##  [3,]     2     0     0     0
##  [4,]     0     0     2     0
##  [5,]    NA     0    NA     2
##  [6,]     1    NA    NA     0
##  [7,]     2     0    NA     2
##  [8,]     2     2     2     2
```

```
##  [9,]     0     0    NA     2
## [10,]     0     1    NA     2
## [11,]     0     0     0    NA
## [12,]     0     1     0    NA
## [13,]     2     0    NA     0
## [14,]     0     2     2     0
## [15,]    NA     2     0     0
## [16,]     0     0    NA     0
## [17,]    NA    NA     0     2
## [18,]     0     0     1    NA
## [19,]     0    NA     1    NA
## [20,]    NA    NA     1     0
## [21,]     1    NA    NA    NA
## [22,]     1     0    NA     0
## [23,]     0    NA     0     0
## [24,]     1    NA    NA     1
## [25,]     1    NA     0     2
## [26,]     1     0     0     0
## [27,]     0     2     0     2
## [28,]     2     0     2     0
## [29,]    NA    NA     1     2
## [30,]     2    NA     0     1
## [31,]     0     2    NA     2
## [32,]     0    NA     1     0
## [33,]    NA     2     1     0
## [34,]     0     0     0     0
## [35,]     1     2     0     1
## [36,]     2    NA     2     2
## [37,]    NA    NA     1    NA
## [38,]     0     2     0     0
## [39,]    NA     0     0    NA
## [40,]     0     0     2    NA
## [41,]     1     2    NA    NA
## [42,]     0     1     2     0
## [43,]    NA     2     1     0
## [44,]     1    NA    NA     2
## [45,]     0    NA    NA     0
## [46,]     0    NA    NA    NA
## [47,]     0     0    NA    NA
## [48,]    NA    NA     0    NA
## [49,]     0     1    NA     1
## [50,]    NA    NA     0     1
```

- We will now learn the `apply` function. This is a handy function that saves writing for loops which should be eschewed in R. Use the apply function to compute a vector whose entries are the standard deviation of each row. Use the apply function to compute a vector whose entries are the standard deviation of each column. Be careful about the NA's! This should be one line.

```
apply(R, 2, sd, na.rm = TRUE)
```

```
##  [1] 0.8243603 0.8553989 0.7800022 0.8181477 0.8337837 0.8725506 0.8698761
##  [8] 0.8518273 0.8243603 0.8600506 0.8700513 0.8431937 0.7601170 0.7828814
## [15] 0.8671182 0.8293400 0.8083372 0.8497042 0.8906612 0.8568215 0.7959984
```

```
## [22] 0.7583370 0.9056473 0.8022905 0.7613390 0.7424692 0.8451543 0.8724011
## [29] 0.8472326 0.8822750 0.9189366 0.8032193 0.7953463 0.8568747 0.7156094
## [36] 0.8669413 0.7310635 0.8335941 0.7593503 0.8929437 0.7355445 0.7249314
## [43] 0.8655918 0.8693637 0.7847758 0.8548905 0.7853242 0.9196554 0.8125775
## [50] 0.9169737
```

- Use the `apply` function to compute a vector whose entries are the count of entries that are 1 or 2 in each column. This should be one line.

```
#count1_2 <- function(x){
#  sum(ifelse(is.na(x) | x == 0, 0, 1))
#}
#apply(R, 2, count1_2)

apply(R>0, 2, sum, na.rm = TRUE)
```

```
##  [1] 15 27 13 15 21 21 21 21 15 18 18 22 21 16 16 17 19 14 21 18 17 17 18
## [24] 19 12 12 20 14 22 17 19 13 20 15 14 20 17 21 13 16 20 17 23 18 23 17
## [47] 16 16 16 17
```

- Use the `split` function to create a list whose keys are the column number and values are the vector of the columns. Look at the last example in the documentation `?split`.

```
split(R, col(R))
```

```
## $`1`
##  [1] NA  1 NA NA  0  1 NA NA  0  0  0 NA  2  2  2  0 NA NA  2  0  1 NA  0
## [24]  0  1  1 NA  0  0  0  0 NA NA  2  1  0 NA  2  1 NA NA  0 NA  2 NA NA
## [47]  0  0 NA  1
##
## $`2`
##  [1] NA  0  0  2  0  2 NA  1  0  2  0  0  2  2  2  2  2  0 NA  2 NA  2  1
## [24]  1  1  2  1 NA  1 NA  0 NA  0  0  0  1  0  1  2  2  0  1  0  1  1  2
## [47] NA  1  0  2
##
## $`3`
##  [1]  0  0  2  0 NA  0 NA  0  0  1 NA  2  2 NA  0 NA  1 NA  1  0  0  0  1
## [24] NA  0 NA  2  1  0  0 NA NA  0  0  0  0 NA  1  0 NA  0  0 NA  2  2 NA
## [47]  1 NA  0  0
##
## $`4`
##  [1]  0  0 NA  0  0  0  0  2 NA  0  1  0 NA  0  0  0  2 NA  0  1  0  0  1
## [24] NA  2  0  1 NA NA  2  0  0 NA  2  1  2  1 NA NA  2  2  0  0  0  1 NA
## [47]  0  0  0  0
##
## $`5`
##  [1] NA  2  0 NA  0  2  0 NA  0  2  1  0 NA NA  0  2  1  1  1  1  0  1  2
## [24]  2 NA NA  1  0 NA  0  0  2  0  0 NA NA  1  0  0 NA NA  1  1  0  2  1
## [47] NA  2  2  0
##
## $`6`
##  [1]  2 NA  0  2  1  1  1 NA NA  0  0 NA  2  2  1 NA NA NA  2  2  0  2  1
```

```
## [24]  2  0 NA NA  0 NA  0  0  2  2 NA  0  2 NA  0  0  0  0 NA  1  1  1  0
## [47] NA  1 NA  2
##
## $`7`
##  [1]  0  0  1 NA  2 NA  2 NA  1  2  0  0 NA  0 NA  2  0  1  2  0 NA  2  2
## [24] NA  2 NA  0  1  2  1  1  1 NA  0 NA NA  0  2  1  0  2 NA  2 NA  0 NA
## [47]  0  1 NA NA
##
## $`8`
##  [1]  2  0  1  0  2  0  0  1 NA  0 NA  0 NA  2  2  2  0  1  0  0  1  1  0
## [24] NA  2  1  2 NA  0 NA  1  2  1  1  0  2  0  0  0 NA  0 NA  0  1  0 NA
## [47]  2  0  0  2
##
## $`9`
##  [1] NA  2  0 NA  0  0  0  2  2 NA NA NA  0 NA NA  0 NA  1  1 NA NA  0  1
## [24]  0 NA NA NA  1 NA  0 NA  2  0  0  0  2  1  0  1  0 NA  2  0  1  1 NA
## [47]  0  2 NA NA
##
## $`10`
##  [1]  0  0 NA  0  1  0  0  0  0  2  2  1 NA NA  2  1 NA  0  2 NA  0  1 NA
## [24]  2 NA  0  0  1  0 NA  0  0  0  0  2  2 NA  2 NA  2 NA  1  0  0  0 NA
## [47]  1  2  1  0
##
## $`11`
##  [1]  1 NA NA  0  2  0  2  0  2  2  0 NA  2  1 NA NA NA  1  0  2  0 NA NA
## [24]  0 NA  0 NA  2 NA NA  1 NA  1 NA  0  0  1  1  2  0  2  0  1 NA  2  0
## [47] NA NA NA NA
##
## $`12`
##  [1]  1  2  0 NA  0 NA  0  0  2  0  1  0  1  0  2  2  0  1  1  1  0  1  0
## [24] NA NA  1  2  2  2  0  2 NA  0  1  2  0 NA NA NA  0  1  0  1  0  2  0
## [47]  0  2  0 NA
##
## $`13`
##  [1]  2  1  1  2 NA  2  1  0  1  0  0  2 NA  0 NA  0 NA  0  0  0  1  2  0
## [24]  0  1 NA NA NA  1  0  0  1 NA  2  2 NA  0  1  1  0  1 NA  1  0  1 NA
## [47] NA NA  1 NA
##
## $`14`
##  [1]  2  0 NA NA  1 NA  1  2 NA  1 NA NA  0  2  0 NA  0  1 NA NA NA NA NA
## [24] NA NA  0  0  0  0 NA  0  2  1  2 NA  0 NA  1  1  2  0  0 NA  0  1  0
## [47]  0  1  1 NA
##
## $`15`
##  [1]  0 NA NA  1  0  1 NA  2  0  2 NA NA NA  0 NA  0  0  0  0  0  1  1 NA
## [24] NA  2  0  0  0  2  1  2  0  1  2 NA  2 NA NA NA NA  2  2 NA NA  0  0
## [47] NA  1  0  0
##
## $`16`
##  [1] NA  2 NA  2  1  0  1  1  0  1 NA  0  0  1  2  0  0  0 NA NA  0 NA  0
## [24]  0  0 NA  1  0  0  0  0  2  2  1 NA  2  0 NA  1  0  0  2  0  0  2  0
## [47]  0  0  0  2
##
## $`17`
```

```
##  [1]  2  1  2  1  0 NA  0  2  0  1 NA NA  1  1 NA NA  0 NA NA NA  0  0 NA
## [24]  2 NA  2  0 NA NA NA  1  0  1  1 NA  0  1 NA  0  2 NA NA  0  2  0  0
## [47]  0  1  1  2
##
## $`18`
##  [1] NA  0  2 NA NA  0  2  0 NA  0  2  0  0  1 NA  1  1  0 NA  1  0  0  0
## [24]  0  0  0  0  0  0  0  0  0 NA NA  2  0  2  0 NA  2  0  0  2  0  1 NA
## [47]  2 NA  2  0
##
## $`19`
##  [1]  0  2  1 NA NA  0 NA NA  2  1  0 NA  2  0  1  0  1  0 NA NA  2  2 NA
## [24] NA NA  2  0  0  2  2  1  2  2  2 NA  0  1  0  2 NA  1  2 NA  0  0  1
## [47] NA NA  0  0
##
## $`20`
##  [1]  0 NA  2  0 NA  0  0  1  0 NA  2 NA  1  0 NA  0 NA  0 NA NA  0  1  0
## [24]  0  1  2  0 NA  0  1  0  0  2 NA  2  2  0 NA  0  1  0  2  0  1  0  2
## [47]  2  1  0  2
##
## $`21`
##  [1]  0  2  1 NA NA  0 NA  1 NA NA  0  0 NA  0 NA  0 NA  0  0  2  0  2  1
## [24] NA  0  0 NA  2 NA  0 NA  1  1  0  1  0  1  0  1 NA  0 NA  2  2 NA  1
## [47]  0  1  0  2
##
## $`22`
##  [1]  1  2  1  0 NA NA  1  1  1  0  0  1  0  0  2 NA  0  0  0  1 NA  2  1
## [24]  0  0  0 NA NA  0 NA  2  0  1 NA  2  0  0  1  0  0 NA NA NA  0  0  0
## [47]  1 NA NA  2
##
## $`23`
##  [1]  1  2 NA NA NA  0  2 NA  2  2  0 NA NA  0  1  0 NA  2 NA NA  0  0 NA
## [24]  0  2 NA NA  2  0  2 NA NA NA NA  1 NA  2 NA  0 NA  0  2 NA  2  1  2
## [47]  0  1  1 NA
##
## $`24`
##  [1]  0  1  1  1  2 NA  1 NA  2  2  2  2  0 NA  0  0  1  0  1  0 NA NA NA
## [24]  0  0  2 NA  0  1  0  1 NA  0 NA  0 NA  2 NA  0  2  0  0  0  0  1  0
## [47]  0 NA  1  1
##
## $`25`
##  [1]  0  0  0  0  2  2  0 NA NA NA NA NA NA  2  0 NA  1 NA  0  2  1  1  2
## [24] NA  0 NA  0 NA  0  0  0  0  0 NA  0  0  1 NA NA  0 NA  1  1  0  0 NA
## [47]  0 NA NA  1
##
## $`26`
##  [1]  2  1  1  0  0  0 NA  2 NA NA  1  0  2  0 NA  0  2  0 NA NA  1  0  1
## [24]  0  0  1  0  0  2 NA  0 NA NA NA NA  0 NA  0  0 NA  0  0  0 NA  0  0
## [47]  0  0  1 NA
##
## $`27`
##  [1]  1 NA NA  0  2  0  2 NA  2  0  1  0  0 NA  0  2 NA  2  2  0  1  2 NA
## [24] NA  2 NA  1 NA  0  0 NA  0 NA  1  0  1  1  0  0  0  2  1  0  2  1 NA
## [47]  1 NA NA  0
##
```

```
## $`28`
##  [1] NA NA  2  0  0  2  1 NA NA  1  2 NA NA  0  1 NA  0  1  2 NA NA NA NA
## [24]  0  1 NA  1 NA NA  0  0  0 NA  0  0  2  2  0  0  0 NA  2  0 NA  0 NA
## [47] NA NA NA  2
##
## $`29`
##  [1]  2  0  2 NA  0  2 NA  1  0 NA NA NA  2 NA NA  0 NA NA NA  2  2  1 NA
## [24]  1  2  0  2  0 NA  2  1  0  1  1 NA  2  0  0 NA  0  1  0  2  1 NA NA
## [47] NA  1  1  2
##
## $`30`
##  [1] NA  0 NA  2  0  1  2  0  0  1  1  2  1  2 NA NA  1  0  0  2 NA  0  0
## [24] NA  1 NA  0 NA  2 NA  0  2  0  2  0 NA  0  1 NA NA NA NA  0  2 NA NA
## [47]  0  2 NA  0
##
## $`31`
##  [1] NA  0  2  2 NA NA  2 NA  2 NA  1  1  2  1  1  2  0  0  1  2  0  2 NA
## [24] NA  0  0 NA NA  0 NA  2  2 NA  0 NA  0  2  0  2  2  1  0  0  0 NA  0
## [47]  0 NA  0  0
##
## $`32`
##  [1] NA  0 NA NA NA NA  0 NA  2  0 NA  0  2  1  0  0  0  1  1  1  0 NA  1
## [24]  1  0  0 NA NA NA  0  0 NA NA  2  2  0  0  0  1 NA  0  2 NA NA  0  2
## [47] NA NA  0 NA
##
## $`33`
##  [1] NA  0  2 NA  1  0  1  0 NA  0  0  1  1  0  0 NA  0  0  1  0  2  1  1
## [24] NA  1 NA  2  0  1 NA NA  0 NA NA  2  1  1 NA NA  2 NA NA  2  2 NA NA
## [47]  0  2 NA  1
##
## $`34`
##  [1]  0  0  2  0  0  2  2  2 NA  0  0 NA  0  0 NA  2  0 NA  0  0  2 NA  0
## [24] NA  0 NA  2  0  1  0  1  1  2 NA NA  0  1  1  0 NA  0  1 NA  2  0 NA
## [47]  0  0  0 NA
##
## $`35`
##  [1]  0  2  0 NA  1 NA  1 NA NA  2  0 NA NA  0  0  0 NA  2  1 NA  0 NA  1
## [24]  0  1 NA  0 NA  2  0  1  0  1  0  0  0 NA NA  0  1  1  0 NA  0 NA  1
## [47] NA  0 NA NA
##
## $`36`
##  [1] NA  0  0  1  2  2 NA  2  0  0 NA NA  2  0 NA  0  0  1  2  2  1 NA  1
## [24]  1  2  1  0  0 NA  0  1  2  1  2  0 NA  0 NA NA  0  0  0 NA  2  0  2
## [47] NA  1 NA NA
##
## $`37`
##  [1] NA  2 NA  0  0 NA  1  0  0  2  1  1 NA  1  0 NA  1 NA  0 NA  0  0  0
## [24]  0  1  1  2 NA  1  0  1  1  0  1  0 NA NA  0 NA  2 NA  0  2 NA  0 NA
## [47]  0  1  0 NA
##
## $`38`
##  [1]  0  2  0  0 NA  0  0  0  0  1  0  1 NA  2 NA NA  1  0  0  0 NA NA  2
## [24] NA  0  0  1  2  2  2  1  2  0  1  1 NA  1 NA NA  1  1  0  2 NA  0  2
## [47]  1  2 NA  0
```

```
##
## $`39`
##  [1]  0 NA NA  0  0  0 NA NA  1  0  0 NA  1  2  0  0 NA NA  0  1 NA  1 NA
## [24] NA  1  0  0  0  0  0 NA  0  0 NA  1 NA NA NA  2 NA NA  0  1  2  0  2
## [47]  1  2  0 NA
##
## $`40`
##  [1]  2  0 NA  0  0 NA  0 NA NA  0  0  1  0  2 NA NA  0  1  2 NA  0  0  0
## [24]  2  0  2  2  2  0  0  0 NA NA  1 NA NA  1 NA  0  2  2 NA  1 NA NA  1
## [47] NA NA  0  2
##
## $`41`
##  [1]  0  2  1 NA  0  2  1  0  0  1 NA  0  2  1  1  0  0  0  1  2  0 NA NA
## [24]  0  2 NA  0  1  0  1  0 NA  1  1  2  1  0  1  0  0 NA  1 NA  0  0  1
## [47]  0  0 NA NA
##
## $`42`
##  [1]  1 NA  0 NA NA  1  2  1  0  1  1 NA  0  1  0  0  1 NA NA  1  0  0  0
## [24]  2  0  0  2 NA  0  1  1  0 NA NA NA  1  0  0  2 NA  0  0  2  0 NA  0
## [47]  1  0 NA  0
##
## $`43`
##  [1]  0 NA  0 NA  1  0 NA NA  0 NA  1  1  0  2  1  0 NA  2  0 NA NA NA  0
## [24]  2  2  2  2  1 NA  1  0 NA  2  2  1  0  0  0  2  2 NA NA  2  0  1  1
## [47]  2  2  0  1
##
## $`44`
##  [1]  1  0  0  0 NA  2 NA  0  1  1 NA NA  1 NA  0  0 NA NA NA  1  2  0  0
## [24] NA NA  1  2  2 NA  2 NA  0  2  0  2 NA NA  2 NA  2  0  1 NA  0  0  0
## [47]  0  2  1  0
##
## $`45`
##  [1]  0  2  0  1 NA  1  2  1  1  1  1  0  1  2 NA NA  1 NA  1  2  0  0  2
## [24]  2  0  0 NA NA  2  0  1  0 NA  2 NA  2 NA  1 NA  1 NA  1 NA  0  0  0
## [47] NA  0 NA  1
##
## $`46`
##  [1]  2 NA  0 NA  2  0  1  0  2  2  1  2  1 NA  0  0  2 NA NA NA  0  0  0
## [24]  1 NA  1  0  0 NA  0  1 NA NA  0 NA NA NA  0 NA  0 NA NA  1  2  0  2
## [47]  0  1  0  2
##
## $`47`
##  [1]  2  0  0 NA  1  0  0  1  1  0  2  0  2  0  0 NA  0  1  0  0  0 NA  0
## [24]  0  0  2  2  1  0  1 NA  0  1 NA NA NA NA  2 NA  0  0  0  1  0  0 NA
## [47]  0  1 NA  2
##
## $`48`
##  [1]  2  0  0 NA NA  2  2  0  0  1 NA  1  0  0  2  2 NA  2 NA  0  0  2  2
## [24]  2  1  0  0  2 NA NA  0 NA NA  0 NA NA NA NA  2  0 NA  0 NA  1  0 NA
## [47]  1 NA NA  0
##
## $`49`
##  [1]  2  0  1  0  0  2  1  0 NA NA  0  2  0  2 NA  1 NA NA  1 NA NA  0  0
## [24]  0  0  2 NA  0  1 NA  0  0 NA NA  0  1  1  2  1 NA NA  0 NA NA  2  1
```

```
## [47] NA NA  0 NA
##
## $`50`
##  [1]  2 NA NA NA  2  0 NA  0  0  0  1  0  0  0  2  0  0 NA NA  2 NA  0  2
## [24]  0 NA  0  0  1  0  2 NA  0  1  2  2  0  2  2  0  0 NA  0 NA  2 NA NA
## [47]  1  0  1  2
```

- In one statement, use the `lapply` function to create a list whose keys are the column number and values are themselves a list with keys: "min" whose value is the minimum of the column, "max" whose value is the maximum of the column, "pct_missing" is the proportion of missingness in the column and "first_NA" whose value is the row number of the first time the NA appears.

```
lapply(split(R, col(R)), function(R){
  list(min = min(R, na.rm = TRUE),
       max = max(R, na.rm = TRUE),
       pct_missing = sum(is.na(R))/n,
       first_NA = min(which(is.na(R)))
  )
})
```

```
## $`1`
## $`1`$min
## [1] 0
##
## $`1`$max
## [1] 2
##
## $`1`$pct_missing
## [1] 0.38
##
## $`1`$first_NA
## [1] 1
##
##
## $`2`
## $`2`$min
## [1] 0
##
## $`2`$max
## [1] 2
##
## $`2`$pct_missing
## [1] 0.16
##
## $`2`$first_NA
## [1] 1
##
##
## $`3`
## $`3`$min
## [1] 0
##
## $`3`$max
```

```
## [1] 2
##
## $`3`$pct_missing
## [1] 0.3
##
## $`3`$first_NA
## [1] 5
##
##
## $`4`
## $`4`$min
## [1] 0
##
## $`4`$max
## [1] 2
##
## $`4`$pct_missing
## [1] 0.22
##
## $`4`$first_NA
## [1] 3
##
##
## $`5`
## $`5`$min
## [1] 0
##
## $`5`$max
## [1] 2
##
## $`5`$pct_missing
## [1] 0.26
##
## $`5`$first_NA
## [1] 1
##
##
## $`6`
## $`6`$min
## [1] 0
##
## $`6`$max
## [1] 2
##
## $`6`$pct_missing
## [1] 0.3
##
## $`6`$first_NA
## [1] 2
##
##
## $`7`
## $`7`$min
## [1] 0
```

```
##
## $`7`$max
## [1] 2
##
## $`7`$pct_missing
## [1] 0.32
##
## $`7`$first_NA
## [1] 4
##
##
## $`8`
## $`8`$min
## [1] 0
##
## $`8`$max
## [1] 2
##
## $`8`$pct_missing
## [1] 0.18
##
## $`8`$first_NA
## [1] 9
##
##
## $`9`
## $`9`$min
## [1] 0
##
## $`9`$max
## [1] 2
##
## $`9`$pct_missing
## [1] 0.38
##
## $`9`$first_NA
## [1] 1
##
##
## $`10`
## $`10`$min
## [1] 0
##
## $`10`$max
## [1] 2
##
## $`10`$pct_missing
## [1] 0.24
##
## $`10`$first_NA
## [1] 3
##
##
## $`11`
```

```
## $`11`$min
## [1] 0
##
## $`11`$max
## [1] 2
##
## $`11`$pct_missing
## [1] 0.38
##
## $`11`$first_NA
## [1] 2
##
##
## $`12`
## $`12`$min
## [1] 0
##
## $`12`$max
## [1] 2
##
## $`12`$pct_missing
## [1] 0.18
##
## $`12`$first_NA
## [1] 4
##
##
## $`13`
## $`13`$min
## [1] 0
##
## $`13`$max
## [1] 2
##
## $`13`$pct_missing
## [1] 0.28
##
## $`13`$first_NA
## [1] 5
##
##
## $`14`
## $`14`$min
## [1] 0
##
## $`14`$max
## [1] 2
##
## $`14`$pct_missing
## [1] 0.38
##
## $`14`$first_NA
## [1] 3
##
```

```
##
## $`15`
## $`15`$min
## [1] 0
##
## $`15`$max
## [1] 2
##
## $`15`$pct_missing
## [1] 0.34
##
## $`15`$first_NA
## [1] 2
##
##
## $`16`
## $`16`$min
## [1] 0
##
## $`16`$max
## [1] 2
##
## $`16`$pct_missing
## [1] 0.18
##
## $`16`$first_NA
## [1] 1
##
##
## $`17`
## $`17`$min
## [1] 0
##
## $`17`$max
## [1] 2
##
## $`17`$pct_missing
## [1] 0.34
##
## $`17`$first_NA
## [1] 6
##
##
## $`18`
## $`18`$min
## [1] 0
##
## $`18`$max
## [1] 2
##
## $`18`$pct_missing
## [1] 0.22
##
## $`18`$first_NA
```

```
## [1] 1
##
##
## $`19`
## $`19`$min
## [1] 0
##
## $`19`$max
## [1] 2
##
## $`19`$pct_missing
## [1] 0.3
##
## $`19`$first_NA
## [1] 4
##
##
## $`20`
## $`20`$min
## [1] 0
##
## $`20`$max
## [1] 2
##
## $`20`$pct_missing
## [1] 0.22
##
## $`20`$first_NA
## [1] 2
##
##
## $`21`
## $`21`$min
## [1] 0
##
## $`21`$max
## [1] 2
##
## $`21`$pct_missing
## [1] 0.3
##
## $`21`$first_NA
## [1] 4
##
##
## $`22`
## $`22`$min
## [1] 0
##
## $`22`$max
## [1] 2
##
## $`22`$pct_missing
## [1] 0.26
```

```
##
## $`22`$first_NA
## [1] 5
##
##
## $`23`
## $`23`$min
## [1] 0
##
## $`23`$max
## [1] 2
##
## $`23`$pct_missing
## [1] 0.42
##
## $`23`$first_NA
## [1] 3
##
##
## $`24`
## $`24`$min
## [1] 0
##
## $`24`$max
## [1] 2
##
## $`24`$pct_missing
## [1] 0.24
##
## $`24`$first_NA
## [1] 6
##
##
## $`25`
## $`25`$min
## [1] 0
##
## $`25`$max
## [1] 2
##
## $`25`$pct_missing
## [1] 0.36
##
## $`25`$first_NA
## [1] 8
##
##
## $`26`
## $`26`$min
## [1] 0
##
## $`26`$max
## [1] 2
##
```

```
## $`26`$pct_missing
## [1] 0.3
##
## $`26`$first_NA
## [1] 7
##
##
## $`27`
## $`27`$min
## [1] 0
##
## $`27`$max
## [1] 2
##
## $`27`$pct_missing
## [1] 0.28
##
## $`27`$first_NA
## [1] 2
##
##
## $`28`
## $`28`$min
## [1] 0
##
## $`28`$max
## [1] 2
##
## $`28`$pct_missing
## [1] 0.42
##
## $`28`$first_NA
## [1] 1
##
##
## $`29`
## $`29`$min
## [1] 0
##
## $`29`$max
## [1] 2
##
## $`29`$pct_missing
## [1] 0.34
##
## $`29`$first_NA
## [1] 4
##
##
## $`30`
## $`30`$min
## [1] 0
##
## $`30`$max
```

```
## [1] 2
##
## $`30`$pct_missing
## [1] 0.34
##
## $`30`$first_NA
## [1] 1
##
##
## $`31`
## $`31`$min
## [1] 0
##
## $`31`$max
## [1] 2
##
## $`31`$pct_missing
## [1] 0.28
##
## $`31`$first_NA
## [1] 1
##
##
## $`32`
## $`32`$min
## [1] 0
##
## $`32`$max
## [1] 2
##
## $`32`$pct_missing
## [1] 0.38
##
## $`32`$first_NA
## [1] 1
##
##
## $`33`
## $`33`$min
## [1] 0
##
## $`33`$max
## [1] 2
##
## $`33`$pct_missing
## [1] 0.34
##
## $`33`$first_NA
## [1] 1
##
##
## $`34`
## $`34`$min
## [1] 0
```

```
##
## $`34`$max
## [1] 2
##
## $`34`$pct_missing
## [1] 0.26
##
## $`34`$first_NA
## [1] 9
##
##
## $`35`
## $`35`$min
## [1] 0
##
## $`35`$max
## [1] 2
##
## $`35`$pct_missing
## [1] 0.36
##
## $`35`$first_NA
## [1] 4
##
##
## $`36`
## $`36`$min
## [1] 0
##
## $`36`$max
## [1] 2
##
## $`36`$pct_missing
## [1] 0.28
##
## $`36`$first_NA
## [1] 1
##
##
## $`37`
## $`37`$min
## [1] 0
##
## $`37`$max
## [1] 2
##
## $`37`$pct_missing
## [1] 0.3
##
## $`37`$first_NA
## [1] 1
##
##
## $`38`
```

```
## $`38`$min
## [1] 0
##
## $`38`$max
## [1] 2
##
## $`38`$pct_missing
## [1] 0.24
##
## $`38`$first_NA
## [1] 5
##
##
## $`39`
## $`39`$min
## [1] 0
##
## $`39`$max
## [1] 2
##
## $`39`$pct_missing
## [1] 0.36
##
## $`39`$first_NA
## [1] 2
##
##
## $`40`
## $`40`$min
## [1] 0
##
## $`40`$max
## [1] 2
##
## $`40`$pct_missing
## [1] 0.34
##
## $`40`$first_NA
## [1] 3
##
##
## $`41`
## $`41`$min
## [1] 0
##
## $`41`$max
## [1] 2
##
## $`41`$pct_missing
## [1] 0.2
##
## $`41`$first_NA
## [1] 4
##
```

```
##
## $`42`
## $`42`$min
## [1] 0
##
## $`42`$max
## [1] 2
##
## $`42`$pct_missing
## [1] 0.26
##
## $`42`$first_NA
## [1] 2
##
##
## $`43`
## $`43`$min
## [1] 0
##
## $`43`$max
## [1] 2
##
## $`43`$pct_missing
## [1] 0.26
##
## $`43`$first_NA
## [1] 2
##
##
## $`44`
## $`44`$min
## [1] 0
##
## $`44`$max
## [1] 2
##
## $`44`$pct_missing
## [1] 0.32
##
## $`44`$first_NA
## [1] 5
##
##
## $`45`
## $`45`$min
## [1] 0
##
## $`45`$max
## [1] 2
##
## $`45`$pct_missing
## [1] 0.28
##
## $`45`$first_NA
```

```
## [1] 5
##
##
## $`46`
## $`46`$min
## [1] 0
##
## $`46`$max
## [1] 2
##
## $`46`$pct_missing
## [1] 0.32
##
## $`46`$first_NA
## [1] 2
##
##
## $`47`
## $`47`$min
## [1] 0
##
## $`47`$max
## [1] 2
##
## $`47`$pct_missing
## [1] 0.22
##
## $`47`$first_NA
## [1] 4
##
##
## $`48`
## $`48`$min
## [1] 0
##
## $`48`$max
## [1] 2
##
## $`48`$pct_missing
## [1] 0.36
##
## $`48`$first_NA
## [1] 4
##
##
## $`49`
## $`49`$min
## [1] 0
##
## $`49`$max
## [1] 2
##
## $`49`$pct_missing
## [1] 0.36
```

```
##
## $`49`$first_NA
## [1] 9
##
##
## $`50`
## $`50`$min
## [1] 0
##
## $`50`$max
## [1] 2
##
## $`50`$pct_missing
## [1] 0.26
##
## $`50`$first_NA
## [1] 2
```

- Create a vector `v` consisting of a sample of 1,000 iid normal realizations with mean -10 and variance 100.

```
v <- rnorm(1000, -10, sqrt(100))
```

- Create a function `my_reverse` which takes as required input a vector and returns the vector in reverse where the first entry is the last entry, etc. No function calls are allowed inside your function otherwise that would defeat the purpose of the exercise! (Yes, there is a base R function that does this called `rev`). Use `head` on `v` and `tail` on `my_reverse(v)` to verify it works.

```
my_reverse <- function(vec){
  vec[length(vec):1]
}

# using `head` on `v` will not be the same as `tail` on `my_reverse(v)` if it reverses properly:
# head(c(1:10)) = {1,2,3,4,5,6},  tail(rev(c(1:10))) = {6,5,4,3,2,1}
my_reverse(head(v)) == tail(my_reverse(v))
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE
```

- Create a function `flip_matrix` which takes as required input a matrix, an argument `dim_to_rev` that returns the matrix with the rows in reverse order or the columns in reverse order depending on the `dim_to_rev` argument. Let the default be the dimension of the matrix that is greater.

```
flip_matrix <- function(X, dim_to_rev = ifelse(dim(X)[1] > dim(X)[2], 1, 2)){
  if (class(X) != "matrix") stop("'X' must be a matrix")
  if (dim_to_rev == 1){
    t(apply(X, 1, my_reverse))
  }
  else{
    apply(X, 2, my_reverse)
  }
}
```

- Find the average of `v` and the standard error of `v`.

```
mean(v)
```

```
## [1] -10.11636
```

```
sd(v)/sqrt(length(v))
```

```
## [1] 0.3144397
```

- Find the 5%ile of `v` and use the `qnorm` function to compute what it theoretically should be. Is the estimate about what is expected by theory?

```
qnorm(0.05, -10, sqrt(100))
```

```
## [1] -26.44854
```

- What is the percentile of `v` that corresponds to the value 0? What should it be theoretically? Is the estimate about what is expected by theory?

```
qnorm(0.85, -10, sqrt(100))
```

```
## [1] 0.3643339
```

- Create a list named `my_list` with keys "A", "B", ... where the entries are arrays of size 1, 2 x 2, 3 x 3 x 3, etc. Fill the array with the numbers 1, 2, 3, etc. Make 8 entries.

```
keys <- LETTERS[1:8]
my_list <- list()
for (i in 1:8){
  my_list[[keys[i]]] = array(1:i, dim = c(rep(i, i)))
}
```

Run the following code:

```
lapply(my_list, object.size)
```

```
## $A
## 224 bytes
##
## $B
## 232 bytes
##
## $C
## 352 bytes
##
## $D
## 1248 bytes
##
```

```
## $E
## 12744 bytes
##
## $F
## 186864 bytes
##
## $G
## 3294416 bytes
##
## $H
## 67109104 bytes
```

Use `?object.size` to read about what these functions do. Then explain the output you see above. For the later arrays, does it make sense given the dimensions of the arrays?

object.size gives you an estimate of the memory allocation attribute of the object, but does not detect if elements in a list are shared. Since it looks at each element as its own, the estimated object size grows exponentially with each new dimension.

Now cleanup the namespace by deleting all stored objects and functions:

```
rm(list = ls())
```

A little about strings

- Use the `strsplit` function and `sample` to put the sentences in the string `lorem` below in random order. You will also need to manipulate the output of `strsplit` which is a list. You may need to learn basic concepts of regular expressions.

```
lorem = "Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi posuere varius volutpat. Morbi

sentences <- paste(unlist(strsplit(lorem, "[.] ")), ".", sep = "")
n <- length(sentences)
vec <- sample(sentences, n)

new <- vec[1]
for (i in 2:n){
  new <- paste(new, vec[i], sep = " ")
}

# pdf_output will not be able to wrap the text
new
```

```
## [1] "Donec vehicula sagittis nisi non semper. Integer dapibus mi lectus, eu posuere arcu ultricies i
```