Unit 1 - Internet Privacy Poll

Problem 1.1

```
poll <- read.csv("AnonymityPoll.csv")
summary(poll)
```

```
##    Internet.Use       Smartphone         Sex           Age
##  Min.   :0.0000   Min.   :0.0000   Female:505   Min.   :18.00
##  1st Qu.:1.0000   1st Qu.:0.0000   Male  :497   1st Qu.:37.00
##  Median :1.0000   Median :1.0000                Median :55.00
##  Mean   :0.7742   Mean   :0.5078                Mean   :52.37
##  3rd Qu.:1.0000   3rd Qu.:1.0000                3rd Qu.:66.00
##  Max.   :1.0000   Max.   :1.0000                Max.   :96.00
##  NA's   :1        NA's   :43                    NA's   :27
##           State            Region   Conservativeness Info.On.Internet
##  California  :103   Midwest  :239   Min.   :1.000    Min.   : 0.000
##  Texas       : 72   Northeast:166   1st Qu.:3.000    1st Qu.: 2.000
##  New York    : 60   South    :359   Median :3.000    Median : 4.000
##  Pennsylvania: 45   West     :238   Mean   :3.277    Mean   : 3.795
##  Florida     : 42                   3rd Qu.:4.000    3rd Qu.: 6.000
##  Ohio        : 38                   Max.   :5.000    Max.   :11.000
##  (Other)     :642                   NA's   :62       NA's   :210
##  Worry.About.Info Privacy.Importance Anonymity.Possible Tried.Masking.Identity
##  Min.   :0.0000   Min.   :  0.00     Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.: 41.43     1st Qu.:0.0000     1st Qu.:0.0000
##  Median :0.0000   Median : 68.75     Median :0.0000     Median :0.0000
##  Mean   :0.4886   Mean   : 62.85     Mean   :0.3692     Mean   :0.1633
##  3rd Qu.:1.0000   3rd Qu.: 88.89     3rd Qu.:1.0000     3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :100.00     Max.   :1.0000     Max.   :1.0000
##  NA's   :212      NA's   :215        NA's   :249        NA's   :218
##  Privacy.Laws.Effective
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2617
##  3rd Qu.:1.0000
##  Max.   :1.0000
##  NA's   :108
```

```
str(poll)
```

```
## 'data.frame':    1002 obs. of  13 variables:
##  $ Internet.Use          : int  1 1 0 1 0 1 1 0 0 1 ...
##  $ Smartphone            : int  0 0 1 0 NA 1 0 0 NA 0 ...
```

```
##  $ Sex                 : Factor w/ 2 levels "Female","Male": 2 2 1 2 1 2 1 1 2 1 ...
##  $ Age                 : int   62 45 70 70 80 49 52 76 75 76 ...
##  $ State               : Factor w/ 49 levels "Alabama","Arizona",..: 20 39 29 10 10 41 21 31 32 32
##  $ Region              : Factor w/ 4 levels "Midwest","Northeast",..: 2 3 2 3 3 3 1 2 3 3 ...
##  $ Conservativeness    : int   4 1 4 4 4 4 3 3 4 4 ...
##  $ Info.On.Internet    : int   0 1 0 3 NA 6 3 NA NA 0 ...
##  $ Worry.About.Info    : int   1 0 0 1 NA 0 1 NA NA 0 ...
##  $ Privacy.Importance  : num   100 0 NA 88.9 NA ...
##  $ Anonymity.Possible  : int   0 1 0 1 NA 1 0 NA NA 1 ...
##  $ Tried.Masking.Identity: int   0 0 0 0 NA 1 0 NA NA 0 ...
##  $ Privacy.Laws.Effective: int   0 1 NA 0 NA 0 1 NA 0 1 ...
```

1002 people participated in the poll.

Problem 1.2

```
table(poll$Smartphone, useNA = "always")
```

```
##
##    0    1 <NA>
##  472  487   43
```

487 interviewees responded that they use a smartphone.
472 interviewees do not use a smartphone.
43 interviewees did not respond to this question.

Problem 1.3

```
table(poll$State, poll$Region)
```

```
##
##                       Midwest Northeast South West
##    Alabama                  0         0    11    0
##    Arizona                  0         0     0   24
##    Arkansas                 0         0    10    0
##    California               0         0     0  103
##    Colorado                 0         0     0   19
##    Connecticut              0         8     0    0
##    Delaware                 0         0     6    0
##    District of Columbia     0         0     2    0
##    Florida                  0         0    42    0
##    Georgia                  0         0    34    0
##    Idaho                    0         0     0    8
##    Illinois                32         0     0    0
##    Indiana                 27         0     0    0
##    Iowa                    14         0     0    0
##    Kansas                  14         0     0    0
##    Kentucky                 0         0    25    0
```

```
##    Louisiana               0         0    17      0
##    Maine                   0         4     0      0
##    Maryland                0         0    18      0
##    Massachusetts           0        19     0      0
##    Michigan               31         0     0      0
##    Minnesota              15         0     0      0
##    Mississippi             0         0    11      0
##    Missouri               26         0     0      0
##    Montana                 0         0     0      5
##    Nebraska               11         0     0      0
##    Nevada                  0         0     0      8
##    New Hampshire           0         7     0      0
##    New Jersey              0        16     0      0
##    New Mexico              0         0     0      5
##    New York                0        60     0      0
##    North Carolina          0         0    32      0
##    North Dakota            5         0     0      0
##    Ohio                   38         0     0      0
##    Oklahoma                0         0    14      0
##    Oregon                  0         0     0     20
##    Pennsylvania            0        45     0      0
##    Rhode Island            0         4     0      0
##    South Carolina          0         0    12      0
##    South Dakota            3         0     0      0
##    Tennessee               0         0    17      0
##    Texas                   0         0    72      0
##    Utah                    0         0     0     11
##    Vermont                 0         3     0      0
##    Virginia                0         0    31      0
##    Washington              0         0     0     28
##    West Virginia           0         0     5      0
##    Wisconsin              23         0     0      0
##    Wyoming                 0         0     0      7
```

Kansas, Missouri and Ohio are in the Midwest region.
Texas is the state in the South with the most interviewees.

Problem 2.1

```r
nrow(subset(poll, Internet.Use == 0 & Smartphone == 0))
```

```
## [1] 186
```

```r
nrow(subset(poll, Internet.Use == 1 & Smartphone == 1))
```

```
## [1] 470
```

```r
nrow(subset(poll, Internet.Use == 1 & Smartphone == 0))
```

```
## [1] 285
```

```r
nrow(subset(poll, Internet.Use == 0 & Smartphone == 1))
```

```
## [1] 17
```

Problem 2.2

```r
table(poll$Internet.Use, useNA = "always")
```

```
##
##    0    1 <NA>
##  226  775    1
```

The answer is 1.

```r
table(poll$Smartphone, useNA = "always")
```

```
##
##    0    1 <NA>
##  472  487   43
```

The answer is 43.

Problem 2.3

```r
limited <- subset(poll, Internet.Use == 1 | Smartphone == 1)
dim(limited)[1]
```

```
## [1] 792
```

Problem 3.1

```r
summary(limited)
```

```
##   Internet.Use      Smartphone         Sex           Age
##  Min.   :0.0000   Min.   :0.0000   Female:392   Min.   :18.00
##  1st Qu.:1.0000   1st Qu.:0.0000   Male  :400   1st Qu.:33.00
##  Median :1.0000   Median :1.0000                Median :51.00
##  Mean   :0.9785   Mean   :0.6308                Mean   :48.57
##  3rd Qu.:1.0000   3rd Qu.:1.0000                3rd Qu.:62.00
##  Max.   :1.0000   Max.   :1.0000                Max.   :93.00
##                   NA's   :20                    NA's   :22
##          State           Region      Conservativeness Info.On.Internet
##  California   : 89   Midwest  :172   Min.   :1.000    Min.   : 0.000
##  Texas        : 57   Northeast:128   1st Qu.:3.000    1st Qu.: 2.000
```

```
##  New York      : 45   South      :288   Median :3.000   Median : 4.000
##  Pennsylvania  : 33   West       :204   Mean   :3.237   Mean   : 3.795
##  Florida       : 32                     3rd Qu.:4.000   3rd Qu.: 6.000
##  North Carolina: 28                     Max.   :5.000   Max.   :11.000
##  (Other)       :508                     NA's   :45
##  Worry.About.Info Privacy.Importance Anonymity.Possible Tried.Masking.Identity
##  Min.   :0.0000   Min.   :  0.00    Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.: 41.43    1st Qu.:0.0000     1st Qu.:0.0000
##  Median :0.0000   Median : 68.75    Median :0.0000     Median :0.0000
##  Mean   :0.4886   Mean   : 62.85    Mean   :0.3692     Mean   :0.1633
##  3rd Qu.:1.0000   3rd Qu.: 88.89    3rd Qu.:1.0000     3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :100.00    Max.   :1.0000     Max.   :1.0000
##  NA's   :2        NA's   :5         NA's   :39         NA's   :8
##  Privacy.Laws.Effective
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2559
##  3rd Qu.:1.0000
##  Max.   :1.0000
##  NA's   :65
```

Problem 3.2

```
mean(limited$Info.On.Internet)
```

```
## [1] 3.795455
```

Problem 3.3

```
table(limited$Info.On.Internet)
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  11
## 105  84  95 101 104  94  67  63  40  18  13   8
```

105, 8

Problem 3.4

```
mean(limited$Worry.About.Info, na.rm = T)
```

```
## [1] 0.4886076
```

Problem 3.5

```r
mean(limited$Anonymity.Possible, na.rm = T)
```

```
## [1] 0.3691899
```

Problem 3.6

```r
mean(limited$Tried.Masking.Identity, na.rm = T)
```

```
## [1] 0.1632653
```

Problem 3.7

```r
mean(limited$Privacy.Laws.Effective, na.rm = T)
```

```
## [1] 0.2558459
```

Problem 4.1

```r
hist(limited$Age)
```

Histogram of limited$Age

People aged about 60 years old

Problem 4.2

```r
max(table(limited$Age, limited$Info.On.Internet))
```
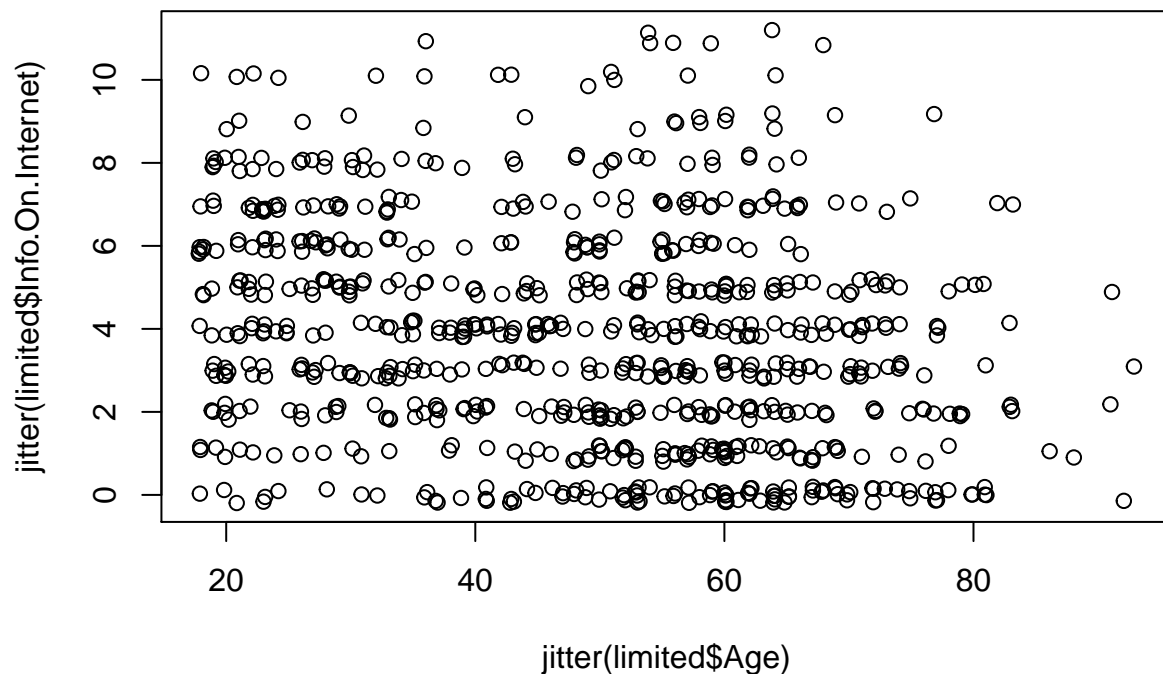
```
## [1] 6
```

Problem 4.3

Jitter adds or subtracts a small amount of random noise to the values passed to it, and two runs will yield different results.

Problem 4.4

```r
plot(jitter(limited$Age), jitter(limited$Info.On.Internet))
```


Older age seems moderately associated with a smaller value for Info.On.Internet

Problem 4.5

```r
tapply(limited$Info.On.Internet, limited$Smartphone, summary)
```

```
## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   3.000   2.923   5.000  11.000
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   2.000   4.000   4.368   6.000  11.000
```

4.368, 2.923

Problem 4.6

```r
tapply(limited$Tried.Masking.Identity, limited$Smartphone, table)
```

```
## $`0`
##
##   0   1
## 248  33
##
## $`1`
##
##   0   1
## 390  93
```

$93/(93+390) = 0.1925$
$33/(33+248) = 0.1174$