

INTERNET PRIVACY POLL

Anuranjan

24 September 2016

The data set has following fields

Internet.Use: A binary variable indicating if the interviewee uses the Internet, at least occasionally (equals 1 if the interviewee uses the Internet, and equals 0 if the interviewee does not use the Internet).

Smartphone: A binary variable indicating if the interviewee has a smartphone (equals 1 if they do have a smartphone, and equals 0 if they don't have a smartphone).

Sex: Male or Female.

Age: Age in years.

State: State of residence of the interviewee.

Region: Census region of the interviewee (Midwest, Northeast, South, or West).

Conservativeness: Self-described level of conservativeness of interviewee, from 1 (very liberal) to 5 (very conservative).

Info.On.Internet:s Number of the following items this interviewee believes to be available on the Internet for others to see: (1) Their email address; (2) Their home address; (3) Their home phone number; (4) Their cell phone number; (5) The employer/company they work for; (6) Their political party or political affiliation; (7) Things they've written that have their name on it; (8) A photo of them; (9) A video of them; (10) Which groups or organizations they belong to; and (11) Their birth date.

Worry>About.Info: A binary variable indicating if the interviewee worries about how much information is available about them on the Internet (equals 1 if they worry, and equals 0 if they don't worry).

Privacy.Importance: A score from 0 (privacy is not too important) to 100 (privacy is very important), which combines the degree to which they find privacy important in the following: (1) The websites they browse; (2) Knowledge of the place they are located when they use the Internet; (3) The content and files they download; (4) The times of day they are online; (5) The applications or programs they use; (6) The searches they perform; (7) The content of their email; (8) The people they exchange email with; and (9) The content of their online chats or hangouts with others.

Anonymity.Possible: A binary variable indicating if the interviewee thinks it's possible to use the Internet anonymously, meaning in such a way that online activities can't be traced back to them (equals 1 if he/she believes you can, and equals 0 if he/she believes you can't).

Tried.Masking.Identity: A binary variable indicating if the interviewee has ever tried to mask his/her identity when using the Internet (equals 1 if he/she has tried to mask his/her identity, and equals 0 if he/she has not tried to mask his/her identity).

Privacy.Laws.Effective: A binary variable indicating if the interviewee believes United States law provides reasonable privacy protection for Internet users (equals 1 if he/she believes it does, and equals 0 if he/she believes it doesn't).

Reading CSV files.

```
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
##  Min.   :0.0000   Min.   : 0.00   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.: 41.43   1st Qu.:0.0000
##  Median :0.0000   Median : 68.75   Median :0.0000
##  Mean   :0.4886   Mean   : 62.85   Mean   :0.3692
##  3rd Qu.:1.0000   3rd Qu.: 88.89   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :100.00   Max.   :1.0000
##  NA's   :212     NA's   :215     NA's   :249
##  Tried.Masking.Identity Privacy.Laws.Effective
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000
##  Mean   :0.1633   Mean   :0.2617
##  3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000
##  NA's   :218     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 3
9 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 ...
```