R Basics for Lecture 1

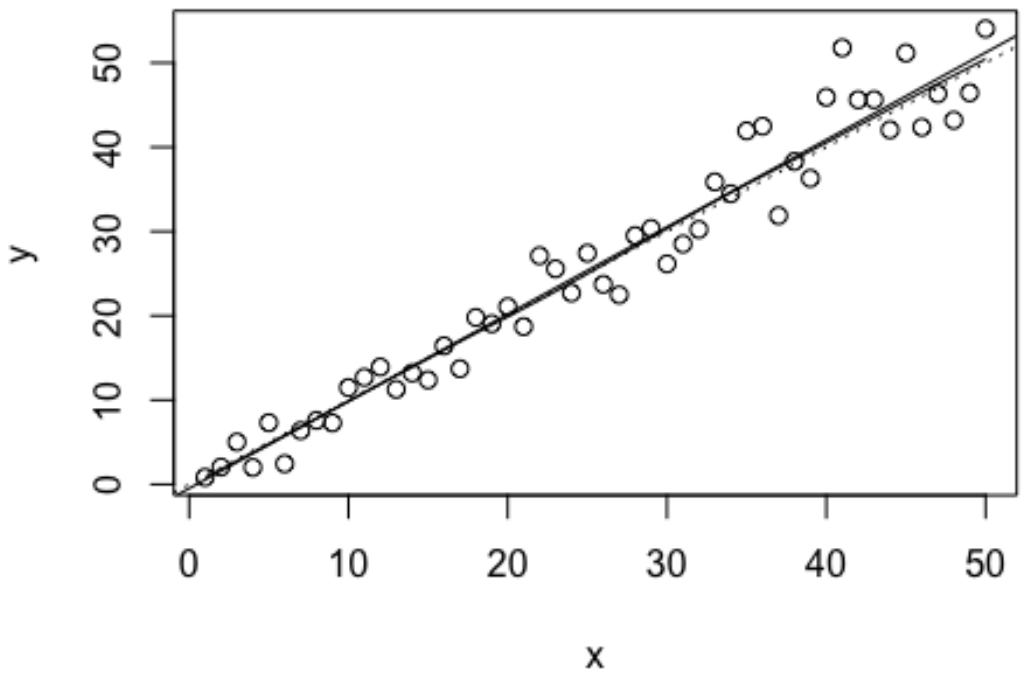
Econ B2000, Statistics and Introduction to Econometrics

```
x <- 1:50
w <- 1 + sqrt(x)/2
example1 <- data.frame(x=x, y= x + rnorm(x)*w)
attach(example1)

fm <- lm(y ~ x)
summary(fm)

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9596 -2.6110 -0.1573  2.0429  9.8693
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.48872    0.97131  -0.503    0.617
## x            1.03417    0.03315  31.196   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.383 on 48 degrees of freedom
## Multiple R-squared:  0.953,  Adjusted R-squared:  0.952
## F-statistic: 973.2 on 1 and 48 DF,  p-value: < 2.2e-16

lrf <- lowess(x, y)
plot(x, y)
lines(x, lrf$y)
abline(0, 1, lty=3)
abline(coef(fm))
```




```
detach()

rm(list = ls(all = TRUE)) # clear workspace

# ^^^^^^

load("acs2017_ny_data.RData")
#glimpse(acs2017_ny) try this later
acs2017_ny[1:10,1:7]
```

```
##    AGE female educ_nohs educ_hs educ_somecoll educ_college educ_advdeg
## 1   72      1         0       0              0            0           1
## 2   72      0         0       0              0            0           1
## 3   31      0         0       0              0            1           0
## 4   28      1         0       0              0            1           0
## 5   54      0         0       0              0            0           1
## 6   45      1         0       1              0            0           0
## 7   84      1         0       0              1            0           0
## 8   71      0         0       0              0            1           0
## 9   68      1         0       0              1            0           0
## 10  37      1         1       0              0            0           0
```

```
attach(acs2017_ny)
```

```
summary(acs2017_ny)

##       AGE            female         educ_nohs        educ_hs
##  Min.   : 0.00   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:22.00   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :42.00   Median :1.0000   Median :0.000   Median :0.0000
##  Mean   :41.57   Mean   :0.5156   Mean   :0.271   Mean   :0.2804
##  3rd Qu.:60.00   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   :95.00   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##
##  educ_somecoll    educ_college     educ_advdeg                  SCHOOL
##  Min.   :0.000   Min.   :0.0000   Min.   :0.000   N/A              :
5569
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000   No, not in school:14
4968
##  Median :0.000   Median :0.0000   Median :0.000   Yes, in school   : 4
6048
##  Mean   :0.173   Mean   :0.1567   Mean   :0.119   Missing          :
0
##  3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.000
##
##                          EDUC
##  Grade 12                  :55119
##  4 years of college        :30802
##  5+ years of college       :23385
##  1 year of college         :19947
##  Nursery school to grade 4:14240
##  2 years of college        :14065
##  (Other)                   :39027
##                                              EDUCD
##  Regular high school diploma                    :35689
##  Bachelor's degree                              :30802
##  1 or more years of college credit, no degree:19947
##  Master's degree                                :17010
##  Associate's degree, type not specified         :14065
##  Some college, but less than 1 year             : 9086
##  (Other)                                        :69986
##                                         DEGFIELD
##  N/A                                        :142398
##  Business                                   :  9802
##  Education Administration and Teaching      :  6708
##  Social Sciences                            :  4836
##  Medical and Health Sciences and Services:  3919
##  Fine Arts                                  :  3491
##  (Other)                                    : 25431
##                                       DEGFIELDD
##  N/A                                       :142398
##  Psychology                                :  2926
##  Business Management and Administration:  2501
```

```
##  Accounting                                :  2284
##  General Education                         :  2238
##  English Language and Literature           :  2202
##  (Other)                                   : 42036
##                                   DEGFIELD2
##  N/A                                      :190425
##  Business                                 :   972
##  Social Sciences                          :   853
##  Education Administration and Teaching:   611
##  Fine Arts                                :   465
##  Communications                           :   352
##  (Other)                                  :  2907
##                                                                  DEGFIELD2D
##  N/A                                                                  :19042
5
##  Psychology                                                           :   28
4
##  Economics                                                            :   26
0
##  Political Science and Government                                     :   24
3
##  Business Management and Administration                               :   21
7
##  French, German, Latin and Other Common Foreign Language Studies:   20
5
##  (Other)                                                              :  495
1
##       PUMA             GQ           OWNERSHP       OWNERSHPD        MOR
TGAGE
##  Min.   : 100   Min.   :1.000   Min.   :0.000   Min.   : 0.00   Min.
:0.000
##  1st Qu.:1500   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:12.00   1st Qu
.:0.000
##  Median :3201   Median :1.000   Median :1.000   Median :13.00   Median
:1.000
##  Mean   :2713   Mean   :1.148   Mean   :1.266   Mean   :14.95   Mean
:1.453
##  3rd Qu.:3902   3rd Qu.:1.000   3rd Qu.:2.000   3rd Qu.:22.00   3rd Qu
.:3.000
##  Max.   :4114   Max.   :5.000   Max.   :2.000   Max.   :22.00   Max.
:4.000
##
##     OWNCOST           RENT         COSTELEC       COSTGAS        COSTW
ATR
##  Min.   :    0   Min.   :   0   Min.   :   0   Min.   :   0   Min.   :
0
##  1st Qu.: 1208   1st Qu.:   0   1st Qu.: 960   1st Qu.: 840   1st Qu.:
320
##  Median : 2891   Median :   0   Median :1560   Median :2400   Median :
1400
```

```
##  Mean   :38582   Mean   : 393   Mean   :2311   Mean   :5032   Mean   :
4836
##  3rd Qu.:99999   3rd Qu.: 630   3rd Qu.:2520   3rd Qu.:9993   3rd Qu.:
9993
##  Max.   :99999   Max.   :3800   Max.   :9997   Max.   :9997   Max.   :
9997
##
##     COSTFUEL       HHINCOME          FOODSTMP        LINGISOL
##  Min.   :   0   Min.   : -11800   Min.   :1.000   Min.   :0.000
##  1st Qu.:9993   1st Qu.:  41600   1st Qu.:1.000   1st Qu.:1.000
##  Median :9993   Median :  81700   Median :1.000   Median :1.000
##  Mean   :7935   Mean   : 114902   Mean   :1.147   Mean   :1.002
##  3rd Qu.:9993   3rd Qu.: 140900   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :9997   Max.   :2030000   Max.   :2.000   Max.   :2.000
##                 NA's   :10630
##      ROOMS           BUILTYR2         UNITSSTR        FUELHEAT
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.00   Min.   :0.000
##  1st Qu.: 4.000   1st Qu.: 1.000   1st Qu.: 3.00   1st Qu.:2.000
##  Median : 6.000   Median : 3.000   Median : 3.00   Median :2.000
##  Mean   : 5.887   Mean   : 3.711   Mean   : 4.39   Mean   :2.959
##  3rd Qu.: 8.000   3rd Qu.: 5.000   3rd Qu.: 6.00   3rd Qu.:4.000
##  Max.   :16.000   Max.   :22.000   Max.   :10.00   Max.   :9.000
##
##       SSMC            FAMSIZE          NCHILD          NCHLT5
##  Min.   :0.00000   Min.   : 1.000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.: 2.000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median : 3.000   Median :0.0000   Median :0.00000
##  Mean   :0.01102   Mean   : 3.087   Mean   :0.5009   Mean   :0.08441
##  3rd Qu.:0.00000   3rd Qu.: 4.000   3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :19.000   Max.   :9.0000   Max.   :5.00000
##
##      RELATE          RELATED           MARST            RACE
RACED
##  Min.   : 1.000   Min.   : 101.0   Min.   :1.000   Min.   :1.00   Min.
:100
##  1st Qu.: 1.000   1st Qu.: 101.0   1st Qu.:1.000   1st Qu.:1.00   1st
Qu.:100
##  Median : 2.000   Median : 201.0   Median :5.000   Median :1.00   Medi
an :100
##  Mean   : 3.307   Mean   : 335.6   Mean   :3.742   Mean   :2.03   Mean
:205
##  3rd Qu.: 3.000   3rd Qu.: 301.0   3rd Qu.:6.000   3rd Qu.:2.00   3rd
Qu.:200
##  Max.   :13.000   Max.   :1301.0   Max.   :6.000   Max.   :9.00   Max.
:990
##
##      HISPAN          HISPAND                  BPL
##  Min.   :0.0000   Min.   :  0.00   New York     :128517
##  1st Qu.:0.0000   1st Qu.:  0.00   West Indies  :  8481
##  Median :0.0000   Median :  0.00   China        :  4964
```

```
##  Mean   :0.4153   Mean   : 44.75   SOUTH AMERICA:  4957
##  3rd Qu.:0.0000   3rd Qu.:  0.00   India        :  3476
##  Max.   :4.0000   Max.   :498.00   Pennsylvania :  3303
##                                    (Other)      : 42887
##                  BPLD                              ANCESTR1
##  New York          :128517   Not Reported            :32021
##  China             :  4116   Italian                 :20577
##  Dominican Republic:  3517   Irish, various subheads,:16388
##  Pennsylvania      :  3303   German                  :12781
##  New Jersey        :  3127   African-American        : 9559
##  Puerto Rico       :  2272   United States           : 8209
##  (Other)           : 51733   (Other)                 :97050
##                                      ANCESTR1D             ANCESTR2
##  Not Reported                             :32021   Not Reported:141487
##  Italian (1990-2000, ACS, PRCS)           :20577   German      :  9476
##  Irish                                    :15651   Irish       :  9238
##  German (1990-2000, ACS/PRCS)             :12605   English     :  4895
##  African-American (1990-2000, ACS, PRCS): 9559   Italian     :  4531
##  United States                            : 8209   Polish      :  3113
##  (Other)                                  :97963   (Other)     : 23845
##                             ANCESTR2D         CITIZEN          YRSUSA1
##  Not Reported                    :141487   Min.   :0.0000   Min.   : 0.0
00
##  German (1990-2000, ACS, PRCS) :  9441   1st Qu.:0.0000   1st Qu.: 0.0
00
##  Irish                         :  8809   Median :0.0000   Median : 0.0
00
##  English                       :  4895   Mean   :0.4793   Mean   : 5.3
77
##  Italian (1990-2000, ACS, PRCS):  4531   3rd Qu.:0.0000   3rd Qu.: 0.0
00
##  Polish                        :  3113   Max.   :3.0000   Max.   :92.0
00
##  (Other)                       : 24309
##     HCOVANY         HCOVPRIV         SEX            EMPSTAT
##  Min.   :1.000   Min.   :1.000   Male  : 95222   Min.   :0.000
##  1st Qu.:2.000   1st Qu.:1.000   Female:101363   1st Qu.:1.000
##  Median :2.000   Median :2.000                   Median :1.000
##  Mean   :1.951   Mean   :1.691                   Mean   :1.514
##  3rd Qu.:2.000   3rd Qu.:2.000                   3rd Qu.:3.000
##  Max.   :2.000   Max.   :2.000                   Max.   :3.000
##
##     EMPSTATD        LABFORCE          OCC              IND
##  Min.   : 0.00   Min.   :0.000   0      : 79987   0      :79987
##  1st Qu.:10.00   1st Qu.:1.000   2310   :  3494   7860   : 9025
##  Median :10.00   Median :2.000   5700   :  3235   8680   : 6354
##  Mean   :15.16   Mean   :1.331   430    :  3025   770    : 6279
##  3rd Qu.:30.00   3rd Qu.:2.000   4720   :  2666   8190   : 5873
##  Max.   :30.00   Max.   :2.000   4760   :  2563   7870   : 4041
##                                  (Other):101615   (Other):85026
```

```
##     CLASSWKR       CLASSWKRD        WKSWORK2        UHRSWORK
##  Min.   :0.000   Min.   : 0.00   Min.   :0.000   Min.   : 0.00
##  1st Qu.:0.000   1st Qu.: 0.00   1st Qu.:0.000   1st Qu.: 0.00
##  Median :2.000   Median :22.00   Median :1.000   Median :12.00
##  Mean   :1.116   Mean   :13.03   Mean   :2.701   Mean   :19.77
##  3rd Qu.:2.000   3rd Qu.:22.00   3rd Qu.:6.000   3rd Qu.:40.00
##  Max.   :2.000   Max.   :29.00   Max.   :6.000   Max.   :99.00
##
##      INCTOT            FTOTINC           INCWAGE          POVERTY
##  Min.   :  -7300   Min.   : -11800   Min.   :     0   Min.   :  0.0
##  1st Qu.:   8000   1st Qu.:  35550   1st Qu.:     0   1st Qu.:159.0
##  Median :  25000   Median :  74000   Median : 10000   Median :351.0
##  Mean   :  45245   Mean   : 107110   Mean   : 33796   Mean   :318.7
##  3rd Qu.:  56500   3rd Qu.: 132438   3rd Qu.: 47000   3rd Qu.:501.0
##  Max.   :1563000   Max.   :2030000   Max.   :638000   Max.   :501.0
##  NA's   :31129     NA's   :10817     NA's   :33427
##     MIGRATE1       MIGRATE1D        MIGPLAC1         MIGCOUNTY1
##  Min.   :0.000   Min.   : 0.00   Min.   :  0.000   Min.   :  0.000
##  1st Qu.:1.000   1st Qu.:10.00   1st Qu.:  0.000   1st Qu.:  0.000
##  Median :1.000   Median :10.00   Median :  0.000   Median :  0.000
##  Mean   :1.122   Mean   :11.51   Mean   :  6.184   Mean   :  4.117
##  3rd Qu.:1.000   3rd Qu.:10.00   3rd Qu.:  0.000   3rd Qu.:  0.000
##  Max.   :4.000   Max.   :40.00   Max.   :900.000   Max.   :810.000
##
##     MIGPUMA1         VETSTAT          VETSTATD         PWPUMA00
##  Min.   :    0   Min.   :0.0000   Min.   : 0.000   Min.   :    0
##  1st Qu.:    0   1st Qu.:1.0000   1st Qu.:11.000   1st Qu.:    0
##  Median :    0   Median :1.0000   Median :11.000   Median :    0
##  Mean   :  277   Mean   :0.8621   Mean   : 9.412   Mean   : 1255
##  3rd Qu.:    0   3rd Qu.:1.0000   3rd Qu.:11.000   3rd Qu.: 3100
##  Max.   :70100   Max.   :2.0000   Max.   :20.000   Max.   :59300
##
##     TRANWORK         TRANTIME         DEPARTS           in_NYC
##  Min.   : 0.000   Min.   :  0.00   Min.   :   0.0   Min.   :0.0000
##  1st Qu.: 0.000   1st Qu.:  0.00   1st Qu.:   0.0   1st Qu.:0.0000
##  Median : 0.000   Median :  0.00   Median :   0.0   Median :0.0000
##  Mean   : 9.725   Mean   : 14.75   Mean   : 373.3   Mean   :0.3615
##  3rd Qu.:10.000   3rd Qu.: 20.00   3rd Qu.: 732.0   3rd Qu.:1.0000
##  Max.   :70.000   Max.   :138.00   Max.   :2345.0   Max.   :1.0000
##
##     in_Bronx        in_Manhattan       in_StatenI       in_Brooklyn
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.000
##  Mean   :0.0538   Mean   :0.04981   Mean   :0.02084   Mean   :0.126
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.000
##
##    in_Queens      in_Westchester      in_Nassau          Hispanic
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
```

```
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.1111   Mean   :0.04413   Mean   :0.07032   Mean   :0.1387
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.0000
##
##     Hisp_Mex          Hisp_PR          Hisp_Cuban          Hisp_DomR
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.000000   Median :0.00000
##  Mean   :0.01626   Mean   :0.0436   Mean   :0.003403   Mean   :0.02827
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.000000   Max.   :1.00000
##
##      white             AfAm           Amindian            Asian
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :1.0000   Median :0.000   Median :0.000000   Median :0.00000
##  Mean   :0.6997   Mean   :0.125   Mean   :0.003779   Mean   :0.08656
##  3rd Qu.:1.0000   3rd Qu.:0.000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.000000   Max.   :1.00000
##
##     race_oth         unmarried        veteran         has_AnyHealthIns
##  Min.   :0.0000   Min.   :0.00   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.00000   1st Qu.:1.0000
##  Median :0.0000   Median :0.00   Median :0.00000   Median :1.0000
##  Mean   :0.1324   Mean   :0.45   Mean   :0.04443   Mean   :0.9513
##  3rd Qu.:0.0000   3rd Qu.:1.00   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.00   Max.   :1.00000   Max.   :1.0000
##
##  has_PvtHealthIns  Commute_car      Commute_bus      Commute_subway
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :1.0000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.6906   Mean   :0.2997   Mean   :0.02162   Mean   :0.07468
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
##
##   Commute_rail     Commute_other     below_povertyline below_150povert
y
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000     Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000     1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.000     Median :0.0000
##  Mean   :0.01332   Mean   :0.05506   Mean   :0.122     Mean   :0.1965
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000     3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000     Max.   :1.0000
##
##  below_200poverty   foodstamps
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000
```

```
##  Median :0.0000   Median :0.0000
##  Mean   :0.2676   Mean   :0.1465
##  3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000
##
```

```
print(NN_obs <- length(AGE))
```

```
## [1] 196585
```

```
summary(AGE[female == 1])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   23.00   44.00   42.72   61.00   95.00
```

```
summary(AGE[!female])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   21.00   40.00   40.35   59.00   95.00
```

```
# here i want to find average ages of men and women
mean(AGE[female == 1])
```

```
## [1] 42.71629
```

```
sd(AGE[female == 1])
```

```
## [1] 23.72012
```

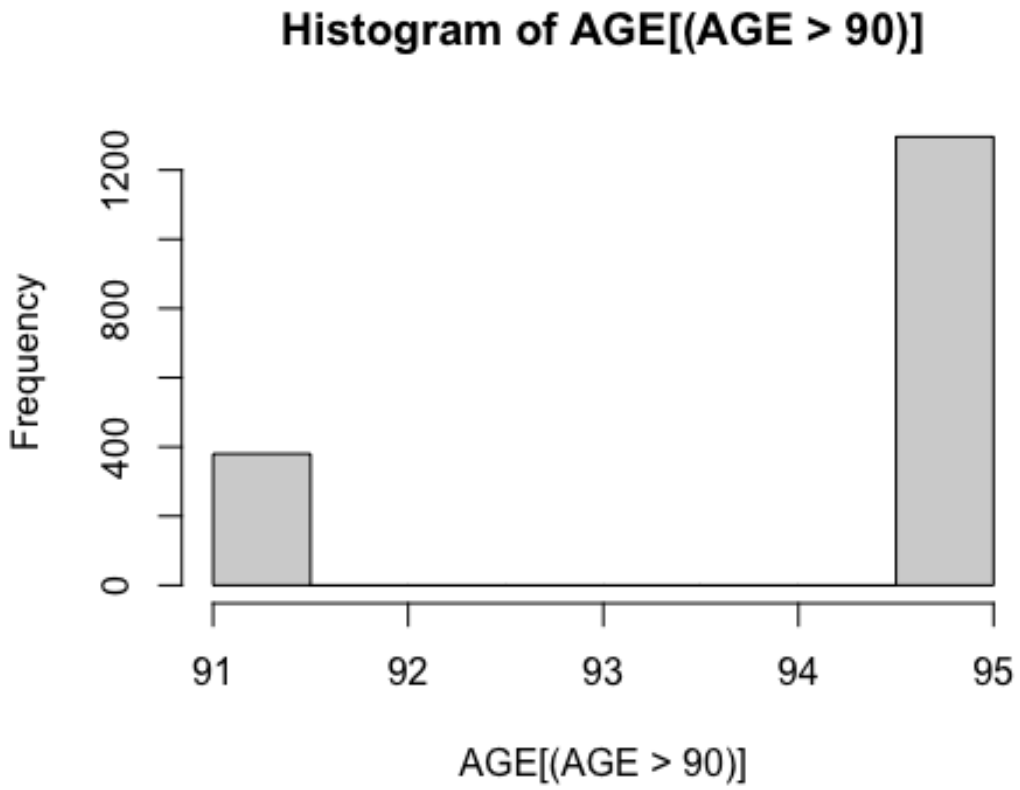
```
mean(AGE[!female])
```

```
## [1] 40.35398
```

```
sd(AGE[!female])
```

```
## [1] 23.1098
```

```
hist(AGE[(AGE > 90)])
```




```
mean(AGE[ (female == 1) & (AGE<90) ])

## [1] 41.98866

str(as.numeric(PUMA))

##  num [1:196585] 902 902 4002 4002 3803 ...
```

PUMA	Neighborhood
3701	NYC-Bronx CD 8–Riverdale, Fieldston & Kingsbridge
3702	NYC-Bronx CD 12–Wakefield, Williamsbridge & Woodlawn
3703	NYC-Bronx CD 10–Co-op City, Pelham Bay & Schuylerville
3704	NYC-Bronx CD 11–Pelham Parkway, Morris Park & Laconia
3705	NYC-Bronx CD 3 & 6–Belmont, Crotona Park East & East Tremont
3706	NYC-Bronx CD 7–Bedford Park, Fordham North & Norwood
3707	NYC-Bronx CD 5–Morris Heights, Fordham South & Mount Hope
3708	NYC-Bronx CD 4–Concourse, Highbridge & Mount Eden
3709	NYC-Bronx CD 9–Castle Hill, Clason Point & Parkchester

3710 NYC-Bronx CD 1 & 2–Hunts Point, Longwood & Melrose
3801 NYC-Manhattan CD 12–Washington Heights, Inwood & Marble Hill
3802 NYC-Manhattan CD 9–Hamilton Heights, Manhattanville & West Harlem
3803 NYC-Manhattan CD 10–Central Harlem
3804 NYC-Manhattan CD 11–East Harlem
3805 NYC-Manhattan CD 8–Upper East Side
3806 NYC-Manhattan CD 7–Upper West Side & West Side
3807 NYC-Manhattan CD 4 & 5–Chelsea, Clinton & Midtown Business District
3808 NYC-Manhattan CD 6–Murray Hill, Gramercy & Stuyvesant Town
3809 NYC-Manhattan CD 3–Chinatown & Lower East Side
3810 NYC-Manhattan CD 1 & 2–Battery Park City, Greenwich Village & Soho
3901 NYC-Staten Island CD 3–Tottenville, Great Kills & Annadale
3902 NYC-Staten Island CD 2–New Springville & South Beach
3903 NYC-Staten Island CD 1–Port Richmond, Stapleton & Mariner's Harbor
4001 NYC-Brooklyn CD 1–Greenpoint & Williamsburg
4002 NYC-Brooklyn CD 4—Bushwick
4003 NYC-Brooklyn CD 3–Bedford-Stuyvesant
4004 NYC-Brooklyn CD 2–Brooklyn Heights & Fort Greene
4005 NYC-Brooklyn CD 6–Park Slope, Carroll Gardens & Red Hook
4006 NYC-Brooklyn CD 8–Crown Heights North & Prospect Heights
4007 NYC-Brooklyn CD 16–Brownsville & Ocean Hill
4008 NYC-Brooklyn CD 5–East New York & Starrett City
4009 NYC-Brooklyn CD 18–Canarsie & Flatlands
4010 NYC-Brooklyn CD 17–East Flatbush, Farragut & Rugby
4011 NYC-Brooklyn CD 9–Crown Heights South, Prospect Lefferts & Wingate
4012 NYC-Brooklyn CD 7–Sunset Park & Windsor Terrace
4013 NYC-Brooklyn CD 10–Bay Ridge & Dyker Heights
4014 NYC-Brooklyn CD 12–Borough Park, Kensington & Ocean Parkway
4015 NYC-Brooklyn CD 14–Flatbush & Midwood
4016 NYC-Brooklyn CD 15–Sheepshead Bay, Gerritsen Beach & Homecrest
4017 NYC-Brooklyn CD 11–Bensonhurst & Bath Beach
4018 NYC-Brooklyn CD 13–Brighton Beach & Coney Island
4101 NYC-Queens CD 1–Astoria & Long Island City
4102 NYC-Queens CD 3–Jackson Heights & North Corona
4103 NYC-Queens CD 7–Flushing, Murray Hill & Whitestone
4104 NYC-Queens CD 11–Bayside, Douglaston & Little Neck

4105 NYC-Queens CD 13–Queens Village, Cambria Heights & Rosedale
4106 NYC-Queens CD 8–Briarwood, Fresh Meadows & Hillcrest
4107 NYC-Queens CD 4–Elmhurst & South Corona
4108 NYC-Queens CD 6–Forest Hills & Rego Park
4109 NYC-Queens CD 2–Sunnyside & Woodside
4110 NYC-Queens CD 5–Ridgewood, Glendale & Middle Village
4111 NYC-Queens CD 9–Richmond Hill & Woodhaven
4112 NYC-Queens CD 12–Jamaica, Hollis & St. Albans
4113 NYC-Queens CD 10–Howard Beach & Ozone Park
4114 NYC-Queens CD 14–Far Rockaway, Breezy Point & Broad Channel

```
PUMA <- as.factor(PUMA)
female <- as.factor(female)

print(levels(female))

## [1] "0" "1"

levels(female) <- c("male","female")
```

Well, ways,

```
educ_indx <- factor((educ_nohs + 2*educ_hs + 3*educ_somecoll + 4*educ_col
lege + 5*educ_advdeg), levels=c(1,2,3,4,5),labels = c("No HS","HS","SmCol
l","Bach","Adv"))

install.packages("tidyverse")
install.packages("plyr")

library(tidyverse)
library(plyr)
levels_n <- read.csv("PUMA_levels.csv")
levels_orig <- levels(PUMA)
levels_new <- join(data.frame(levels_orig),data.frame(levels_n))
levels(PUMA) <- levels_new$New_Level

summary(female)

##   male female
##  95222 101363

summary(PUMA)

##                   NYC-Bronx CD 8--Riverdale, Fieldston & Kingsbridge
##                                                                  936
##                NYC-Bronx CD 12--Wakefield, Williamsbridge & Woodlawn
##                                                                 1092
##              NYC-Bronx CD 10--Co-op City, Pelham Bay & Schuylerville
##                                                                  767
##               NYC-Bronx CD 11--Pelham Parkway, Morris Park & Laconia
```

```
##                                                                 1115
##        NYC-Bronx CD 3 & 6--Belmont, Crotona Park East & East Tremont
##                                                                 1311
##                NYC-Bronx CD 7--Bedford Park, Fordham North & Norwood
##                                                                  854
##           NYC-Bronx CD 5--Morris Heights, Fordham South & Mount Hope
##                                                                 1112
##                    NYC-Bronx CD 4--Concourse, Highbridge & Mount Eden
##                                                                  917
##              NYC-Bronx CD 9--Castle Hill, Clason Point & Parkchester
##                                                                 1307
##                  NYC-Bronx CD 1 & 2--Hunts Point, Longwood & Melrose
##                                                                 1166
##        NYC-Manhattan CD 12--Washington Heights, Inwood & Marble Hill
##                                                                 1238
##   NYC-Manhattan CD 9--Hamilton Heights, Manhattanville & West Harlem
##                                                                  872
##                                    NYC-Manhattan CD 10--Central Harlem
##                                                                  897
##                                       NYC-Manhattan CD 11--East Harlem
##                                                                  769
##                                    NYC-Manhattan CD 8--Upper East Side
##                                                                 1167
##                       NYC-Manhattan CD 7--Upper West Side & West Side
##                                                                  949
## NYC-Manhattan CD 4 & 5--Chelsea, Clinton & Midtown Business District
##                                                                  944
##          NYC-Manhattan CD 6--Murray Hill, Gramercy & Stuyvesant Town
##                                                                  758
##                       NYC-Manhattan CD 3--Chinatown & Lower East Side
##                                                                 1062
##  NYC-Manhattan CD 1 & 2--Battery Park City, Greenwich Village & Soho
##                                                                 1136
##          NYC-Staten Island CD 3--Tottenville, Great Kills & Annadale
##                                                                 1303
##                NYC-Staten Island CD 2--New Springville & South Beach
##                                                                 1173
##  NYC-Staten Island CD 1--Port Richmond, Stapleton & Mariner's Harbor
##                                                                 1621
##                           NYC-Brooklyn CD 1--Greenpoint & Williamsburg
##                                                                 1293
##                                          NYC-Brooklyn CD 4--Bushwick
##                                                                 1060
##                                   NYC-Brooklyn CD 3--Bedford-Stuyvesant
##                                                                 1082
##                     NYC-Brooklyn CD 2--Brooklyn Heights & Fort Greene
##                                                                 1320
##            NYC-Brooklyn CD 6--Park Slope, Carroll Gardens & Red Hook
##                                                                 1168
##            NYC-Brooklyn CD 8--Crown Heights North & Prospect Heights
```

```
##                                                                 1077
##                         NYC-Brooklyn CD 16--Brownsville & Ocean Hill
##                                                                  904
##                     NYC-Brooklyn CD 5--East New York & Starrett City
##                                                                 1321
##                              NYC-Brooklyn CD 18--Canarsie & Flatlands
##                                                                 2422
##                   NYC-Brooklyn CD 17--East Flatbush, Farragut & Rugby
##                                                                 1250
##  NYC-Brooklyn CD 9--Crown Heights South, Prospect Lefferts & Wingate
##                                                                  818
##                     NYC-Brooklyn CD 7--Sunset Park & Windsor Terrace
##                                                                 1291
##                        NYC-Brooklyn CD 10--Bay Ridge & Dyker Heights
##                                                                 1519
##         NYC-Brooklyn CD 12--Borough Park, Kensington & Ocean Parkway
##                                                                 1698
##                                  NYC-Brooklyn CD 14--Flatbush & Midwood
##                                                                 1479
##      NYC-Brooklyn CD 15--Sheepshead Bay, Gerritsen Beach & Homecrest
##                                                                 1903
##                           NYC-Brooklyn CD 11--Bensonhurst & Bath Beach
##                                                                 2234
##                     NYC-Brooklyn CD 13--Brighton Beach & Coney Island
##                                                                  925
##                           NYC-Queens CD 1--Astoria & Long Island City
##                                                                 1748
##                        NYC-Queens CD 3--Jackson Heights & North Corona
##                                                                 1316
##                   NYC-Queens CD 7--Flushing, Murray Hill & Whitestone
##                                                                 2290
##                   NYC-Queens CD 11--Bayside, Douglaston & Little Neck
##                                                                 1344
##         NYC-Queens CD 13--Queens Village, Cambria Heights & Rosedale
##                                                                 2148
##                 NYC-Queens CD 8--Briarwood, Fresh Meadows & Hillcrest
##                                                                 1393
##                                NYC-Queens CD 4--Elmhurst & South Corona
##                                                                  973
##                              NYC-Queens CD 6--Forest Hills & Rego Park
##                                                                 1041
##                                   NYC-Queens CD 2--Sunnyside & Woodside
##                                                                 1158
##                 NYC-Queens CD 5--Ridgewood, Glendale & Middle Village
##                                                                 2040
##                              NYC-Queens CD 9--Richmond Hill & Woodhaven
##                                                                 1694
##                         NYC-Queens CD 12--Jamaica, Hollis & St. Albans
##                                                                 2438
##                              NYC-Queens CD 10--Howard Beach & Ozone Park
```

```
##                                                                    1304
##         NYC-Queens CD 14--Far Rockaway, Breezy Point & Broad Channel
##                                                                     954
##                                                                    NA's
##                                                                  125514
```

```
summary(educ_indx)
```

```
##  No HS     HS SmColl   Bach    Adv
##  53267  55119  34012  30802  23385
```

```
ddply(acs2017_ny, .(PUMA), summarize, mean = round(mean(AGE), 2), sd = ro
und(sd(AGE), 2))
```

```
##     PUMA  mean    sd
## 1    100 41.70 23.85
## 2    200 43.47 23.45
## 3    300 44.88 23.60
## 4    401 45.12 24.28
## 5    402 42.23 24.18
## 6    403 43.22 23.66
## 7    500 40.01 23.73
## 8    600 39.77 23.29
## 9    701 36.64 22.44
## 10   702 42.13 23.02
## 11   703 43.70 24.11
## 12   704 44.71 23.83
## 13   800 43.93 23.62
## 14   901 45.35 24.04
## 15   902 40.74 22.29
## 16   903 38.70 22.42
## 17   904 43.70 23.84
## 18   905 40.84 22.75
## 19   906 41.48 24.26
## 20  1000 42.77 23.43
## 21  1101 44.33 24.22
## 22  1102 42.49 23.23
## 23  1201 44.28 24.74
## 24  1202 42.72 24.68
## 25  1203 43.80 23.84
## 26  1204 43.57 23.84
## 27  1205 40.35 23.50
## 28  1206 37.94 22.96
## 29  1207 45.42 23.92
## 30  1300 43.73 23.35
## 31  1400 43.80 24.01
## 32  1500 40.94 23.49
## 33  1600 43.30 24.68
## 34  1700 43.52 23.86
## 35  1801 42.63 23.60
## 36  1802 43.08 23.23
```

```
## 37  1900 42.81 23.57
## 38  2001 37.81 22.86
## 39  2002 43.45 23.90
## 40  2100 46.88 23.54
## 41  2201 41.39 23.92
## 42  2202 44.64 23.67
## 43  2203 44.31 24.08
## 44  2300 37.97 22.08
## 45  2401 45.15 23.39
## 46  2402 42.78 24.42
## 47  2500 42.73 24.33
## 48  2600 43.71 24.47
## 49  2701 44.76 23.71
## 50  2702 42.69 23.16
## 51  2801 43.81 23.74
## 52  2802 43.19 23.91
## 53  2901 41.72 23.03
## 54  2902 42.99 22.82
## 55  2903 34.24 23.39
## 56  3001 44.04 23.76
## 57  3002 42.34 24.24
## 58  3003 32.23 24.23
## 59  3101 43.50 22.72
## 60  3102 43.58 23.51
## 61  3103 43.50 23.89
## 62  3104 39.12 23.54
## 63  3105 42.69 24.40
## 64  3106 42.68 23.85
## 65  3107 42.05 23.57
## 66  3201 43.25 25.02
## 67  3202 44.08 24.19
## 68  3203 43.40 24.06
## 69  3204 43.77 23.56
## 70  3205 42.64 23.83
## 71  3206 40.57 23.55
## 72  3207 42.79 24.08
## 73  3208 43.68 23.96
## 74  3209 43.11 23.37
## 75  3210 42.34 23.58
## 76  3211 41.49 22.93
## 77  3212 44.51 25.12
## 78  3301 45.37 24.21
## 79  3302 43.40 24.24
## 80  3303 43.01 24.03
## 81  3304 41.35 23.85
## 82  3305 48.60 24.31
## 83  3306 43.79 22.92
## 84  3307 41.86 22.68
## 85  3308 41.31 23.04
## 86  3309 43.69 23.07
```

```
## 87  3310 39.30 22.38
## 88  3311 40.25 23.23
## 89  3312 40.43 22.31
## 90  3313 40.90 23.82
## 91  3701 43.12 25.58
## 92  3702 40.22 22.96
## 93  3703 43.63 24.07
## 94  3704 42.05 24.57
## 95  3705 34.78 22.47
## 96  3706 35.15 22.40
## 97  3707 33.70 22.15
## 98  3708 35.25 22.01
## 99  3709 38.88 23.99
## 100 3710 35.34 22.09
## 101 3801 40.55 22.58
## 102 3802 35.62 20.80
## 103 3803 39.45 21.16
## 104 3804 38.39 21.37
## 105 3805 43.53 23.63
## 106 3806 42.44 22.85
## 107 3807 40.20 19.30
## 108 3808 40.66 21.37
## 109 3809 40.98 22.18
## 110 3810 39.03 20.90
## 111 3901 42.89 23.76
## 112 3902 41.08 23.88
## 113 3903 40.75 22.72
## 114 4001 35.39 20.42
## 115 4002 34.12 19.37
## 116 4003 36.03 20.78
## 117 4004 36.75 20.23
## 118 4005 36.84 20.31
## 119 4006 38.50 20.69
## 120 4007 39.63 21.48
## 121 4008 36.65 22.17
## 122 4009 41.51 23.14
## 123 4010 42.14 23.08
## 124 4011 39.77 22.98
## 125 4012 36.75 21.46
## 126 4013 42.91 22.97
## 127 4014 35.35 24.37
## 128 4015 38.65 23.33
## 129 4016 42.18 24.11
## 130 4017 39.67 22.74
## 131 4018 45.54 24.77
## 132 4101 38.65 20.21
## 133 4102 38.72 22.55
## 134 4103 44.60 23.11
## 135 4104 45.76 23.24
## 136 4105 41.99 23.29
```

```
## 137 4106 40.17 23.88
## 138 4107 40.09 21.87
## 139 4108 42.64 23.42
## 140 4109 41.20 20.83
## 141 4110 40.21 22.13
## 142 4111 40.69 21.64
## 143 4112 40.37 22.80
## 144 4113 39.78 22.77
## 145 4114 39.47 24.25
```

```
dat_use1 <- subset(acs2017_ny,((INCWAGE > 0) & in_NYC))
ddply(dat_use1, .(PUMA), summarize, inc90 = quantile(INCWAGE,probs = 0.9)
, inc10 = quantile(INCWAGE,probs = 0.1), n_obs = length(INCWAGE))
```

```
##    PUMA  inc90 inc10 n_obs
## 1  3701 120000  3220   424
## 2  3702  85000  6700   541
## 3  3703 100500  3750   366
## 4  3704  90000  6980   510
## 5  3705  52000  3000   537
## 6  3706  63200  5940   359
## 7  3707  60000  4000   439
## 8  3708  62000  6000   376
## 9  3709  78800  5220   503
## 10 3710  55000  3580   420
## 11 3801 100000  5000   670
## 12 3802 120000  3000   399
## 13 3803 130000  6000   478
## 14 3804 120000  7000   368
## 15 3805 300000 17900   636
## 16 3806 326000  7860   509
## 17 3807 268000 10000   635
## 18 3808 300000 20560   460
## 19 3809 140000  5000   515
## 20 3810 300000  6000   695
## 21 3901 127000  6220   617
## 22 3902 125000  8000   524
## 23 3903 100000  7100   771
## 24 4001 149500 10000   736
## 25 4002  82000  9000   581
## 26 4003 110000  7200   557
## 27 4004 166000  7000   786
## 28 4005 200000 12000   681
## 29 4006 114000  8740   585
## 30 4007  79000  4800   361
## 31 4008  73000  6000   549
## 32 4009 100000  9600  1178
## 33 4010  80200  8360   610
## 34 4011  95400  7000   407
## 35 4012 102200  6880   625
```

```
## 36 4013 124000  7440   773
## 37 4014  90000  5590   654
## 38 4015 100000  7450   710
## 39 4016 101200  6000   899
## 40 4017  97000  7200  1070
## 41 4018 100000  5000   368
## 42 4101 104000  9600  1041
## 43 4102  82400  8000   624
## 44 4103 100000  7180  1107
## 45 4104 110000  8000   661
## 46 4105 100000  7000  1080
## 47 4106 102000  8000   641
## 48 4107  70000  8000   499
## 49 4108 140000 10000   563
## 50 4109 129600 11600   655
## 51 4110 100000 10000  1049
## 52 4111  90000  8680   865
## 53 4112  84800  7260  1213
## 54 4113  93000  6000   625
## 55 4114 108300  6700   378

table(educ_indx,female)

##          female
## educ_indx  male female
##    No HS  27180  26087
##    HS     27309  27810
##    SmColl 15847  18165
##    Bach   14632  16170
##    Adv    10254  13131

xtabs(~educ_indx + female)

##          female
## educ_indx  male female
##    No HS  27180  26087
##    HS     27309  27810
##    SmColl 15847  18165
##    Bach   14632  16170
##    Adv    10254  13131
```

Want proportions instead of counts?

```
prop.table(table(educ_indx,female))

##          female
## educ_indx       male     female
##    No HS  0.13826080 0.13270087
##    HS     0.13891701 0.14146552
##    SmColl 0.08061144 0.09240278
##    Bach   0.07443091 0.08225450
##    Adv    0.05216064 0.06679553
```

```r
mean(educ_nohs[(AGE >= 25)&(AGE <= 55)])
```

```
## [1] 0.08354656
```

```r
mean(educ_hs[(AGE >= 25)&(AGE <= 55)])
```

```
## [1] 0.2974594
```

```r
mean(educ_somecoll[(AGE >= 25)&(AGE <= 55)])
```

```
## [1] 0.2057843
```

```r
mean(educ_college[(AGE >= 25)&(AGE <= 55)])
```

```
## [1] 0.2383112
```

```r
mean(educ_advdeg[(AGE >= 25)&(AGE <= 55)])
```

```
## [1] 0.1748986
```

```r
# alternatively
restrict1 <- as.logical((AGE >= 25)&(AGE <= 55))
dat_age_primeage <- subset(acs2017_ny, restrict1)

detach()
attach(dat_age_primeage)

mean(educ_nohs)
```

```
## [1] 0.08354656
```

```r
mean(educ_hs)
```

```
## [1] 0.2974594
```

```r
mean(educ_somecoll)
```

```
## [1] 0.2057843
```

```r
mean(educ_college)
```

```
## [1] 0.2383112
```

```r
mean(educ_advdeg)
```

```
## [1] 0.1748986
```

```r
detach()
```