MEM_second

chenfan

2024-11-30

R Markdown

Question #1:BigBangTheory. (Attached Data: BigBangTheory)

The Big Bang Theory, a situation comedy featuring Johnny Galecki, Jim Parsons, and Kaley Cuoco-Sweeting, is one of the most-watched programs on network television. The first two episodes for the 2011–2012 season premiered on September 22, 2011; the first episode attracted 14.1 million viewers and the second episode attracted 14.7 million viewers. The attached data file BigBangTheory shows the number of viewers in millions for the first 21 episodes of the 2011–2012 season (the Big Bang theory website, April 17, 2012). a. Compute the minimum and the maximum number of viewers. b. Compute the mean, median, and mode. c. Compute the first and third quartiles. d. has viewership grown or declined over the 2011–2012 season? Discuss

```
#getwd()
data <- read.csv("data/BigBangTheory.csv", header = TRUE, sep = ",")
data
```

```
##              Air.Date Viewers..millions.
## 1  September 22, 2011               14.1
## 2  September 22, 2011               14.7
## 3  September 29, 2011               14.6
## 4     October 6, 2011               13.6
## 5    October 13, 2011               13.6
## 6    October 20, 2011               14.9
## 7    October 27, 2011               14.5
## 8    November 3, 2011               16.0
## 9   November 10, 2011               15.9
## 10  November 17, 2011               15.1
## 11   December 8, 2011               14.0
## 12   January 12, 2012               16.1
## 13   January 19, 2012               15.8
## 14   January 26, 2012               16.1
## 15   February 2, 2012               16.5
## 16   February 9, 2012               16.2
## 17  February 16, 2012               15.7
## 18  February 23, 2012               16.2
## 19      March 8, 2012               15.0
## 20     March 29, 2012               14.0
## 21      April 5, 2012               13.3
```

```
max(data$Viewers..millions.)
```

```
## [1] 16.5
```

```
min(data$Viewers..millions.)
```

```
## [1] 13.3
```

```
mean(data$Viewers..millions.)
```

```
## [1] 15.04286
```

```
median(data$Viewers..millions.)
```

```
## [1] 15
```

```
as.numeric(names(which.max(table(data$Viewers..millions.))))
```

```
## [1] 13.6
```

```
quantile(data$Viewers, 0.25, na.rm = TRUE)
```

```
##  25%
## 14.1
```

```
quantile(data$Viewers, 0.75, na.rm = TRUE)
```

```
## 75%
##  16
```

```
model <- lm(Viewers..millions. ~ Air.Date, data = data)
#coef(model)
summary(model)
```

```
##
## Call:
## lm(formula = Viewers..millions. ~ Air.Date, data = data)
##
## Residuals:
##          1          2          3          4          5          6          7
## -3.000e-01  3.000e-01 -2.082e-17  4.163e-17 -2.082e-17  3.469e-17  3.469e-17
##          8          9         10         11         12         13         14
## -1.388e-17 -6.939e-18  0.000e+00  1.735e-16  2.776e-17  6.939e-18  1.388e-17
##         15         16         17         18         19         20         21
##  1.388e-17 -2.082e-17 -6.939e-18  6.939e-18 -1.804e-16 -1.388e-17  1.249e-16
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 13.3000     0.4243  31.348   0.0203 *
## Air.DateDecember 8, 2011     0.7000     0.6000   1.167   0.4511
## Air.DateFebruary 16, 2012    2.4000     0.6000   4.000   0.1560
## Air.DateFebruary 2, 2012     3.2000     0.6000   5.333   0.1180
## Air.DateFebruary 23, 2012    2.9000     0.6000   4.833   0.1299
## Air.DateFebruary 9, 2012     2.9000     0.6000   4.833   0.1299
## Air.DateJanuary 12, 2012     2.8000     0.6000   4.667   0.1344
## Air.DateJanuary 19, 2012     2.5000     0.6000   4.167   0.1500
## Air.DateJanuary 26, 2012     2.8000     0.6000   4.667   0.1344
## Air.DateMarch 29, 2012       0.7000     0.6000   1.167   0.4511
## Air.DateMarch 8, 2012        1.7000     0.6000   2.833   0.2160
## Air.DateNovember 10, 2011    2.6000     0.6000   4.333   0.1444
## Air.DateNovember 17, 2011    1.8000     0.6000   3.000   0.2048
## Air.DateNovember 3, 2011     2.7000     0.6000   4.500   0.1392
## Air.DateOctober 13, 2011     0.3000     0.6000   0.500   0.7048
## Air.DateOctober 20, 2011     1.6000     0.6000   2.667   0.2284
## Air.DateOctober 27, 2011     1.2000     0.6000   2.000   0.2952
## Air.DateOctober 6, 2011      0.3000     0.6000   0.500   0.7048
## Air.DateSeptember 22, 2011   1.1000     0.5196   2.117   0.2809
## Air.DateSeptember 29, 2011   1.3000     0.6000   2.167   0.2753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4243 on 1 degrees of freedom
## Multiple R-squared:  0.9913, Adjusted R-squared:  0.8252
## F-statistic: 5.968 on 19 and 1 DF,  p-value: 0.3131
```

```
#p<0.05  no contact
```

Question #2: NBAPlayerPts. (Attached Data: NBAPlayerPts)

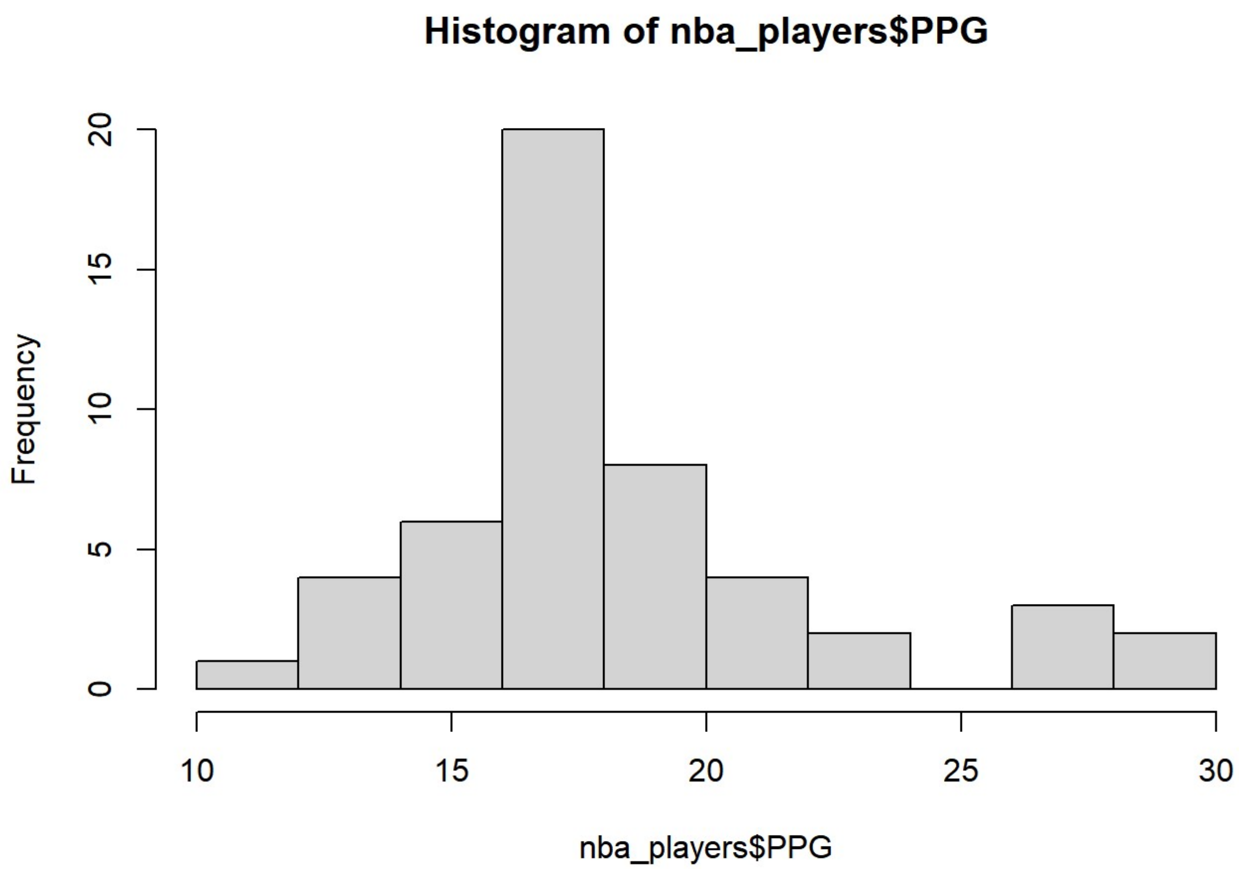
CbSSports.com developed the Total Player Rating system to rate players in the National Basketball Association (NBA) based on various offensive and defensive statistics. The attached data file NBAPlayerPts shows the average number of points scored per game (PPG) for 50 players with the highest ratings for a portion of the 2012–2013 NBA season (CbSSports.com website, February 25, 2013). Use classes starting at 10 and ending at 30 in increments of 2 for PPG in the following. a. Show the frequency distribution. b.

Show the relative frequency distribution. c. Show the cumulative percent frequency distribution. d. Develop a histogram for the average number of points scored per game. e. Do the data appear to be skewed? Explain. f. What percentage of the players averaged at least 20 points per game?

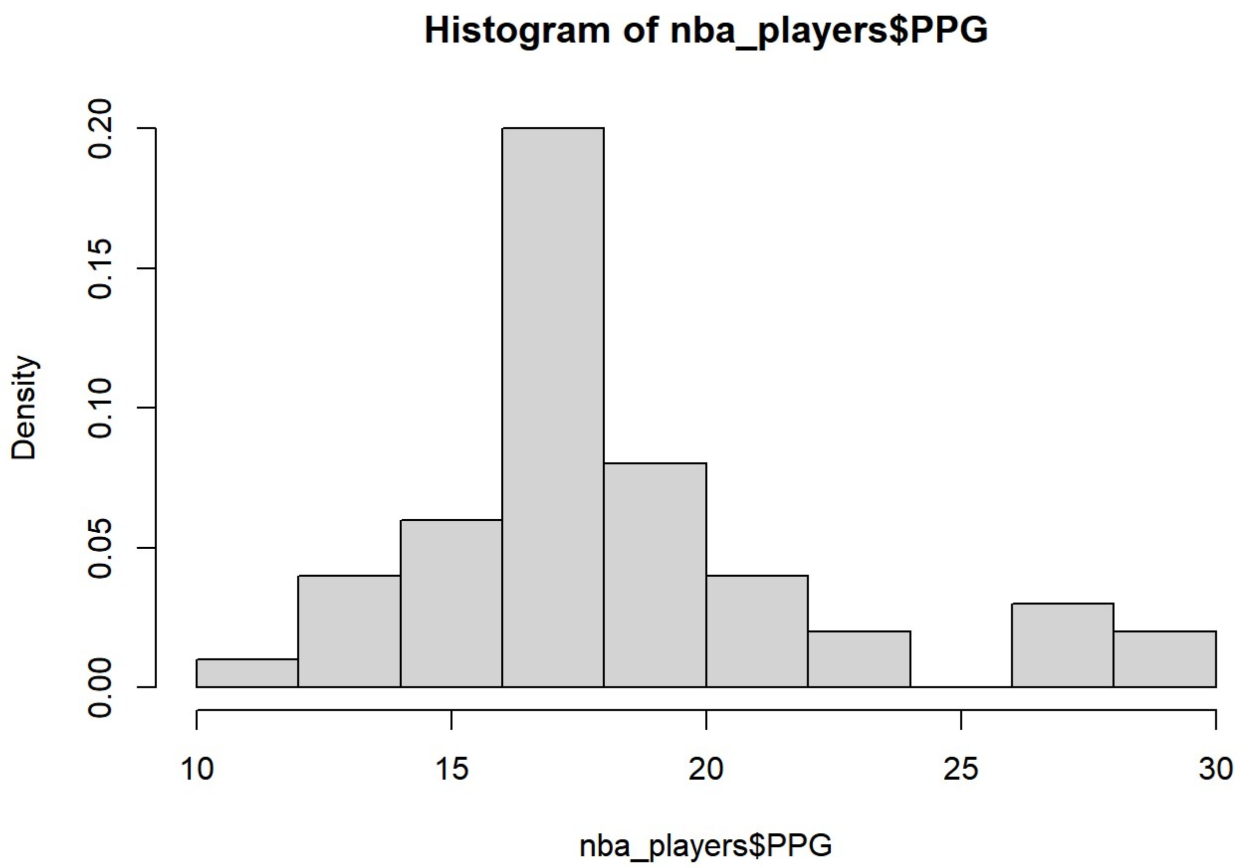
```
nba_players <- read.csv("data/NBAPlayerPts.csv", header = TRUE, sep = ",")
nba_players
```

```
##    Rank                         Player  PPG
## 1     1              LeBron James, MIA 27.0
## 2     2              Kevin Durant, OKC 28.8
## 3     3              James Harden, HOU 26.4
## 4     4                Kobe Bryant, LAL 27.1
## 5     5         Russell Westbrook, OKC 22.9
## 6     6            Carmelo Anthony, NY 28.4
## 7     7                  David Lee, GS 19.2
## 8     8              Stephen Curry, GS 21.0
## 9     9         LaMarcus Aldridge, POR 20.8
## 10   10               Paul George, IND 17.6
## 11   11                Tony Parker, SA 21.1
## 12   12              Jrue Holiday, PHI 19.2
## 13   13               Dwyane Wade, MIA 21.2
## 14   14             Nicolas Batum, POR 15.5
## 15   15                Josh Smith, ATL 17.2
## 16   16                Al Horford, ATL 16.7
## 17   17              Al Jefferson, UTA 17.6
## 18   18             Blake Griffin, LAC 18.5
## 19   19               Paul Pierce, BOS 18.3
## 20   20 Damian Lillard, POR (Rookie) 18.3
## 21   21              Kyrie Irving, CLE 23.3
## 22   22             Dwight Howard, LAL 16.4
## 23   23          Brandon Jennings, MIL 18.9
## 24   24                 Luol Deng, CHI 16.5
## 25   25            Deron Williams, BKN 17.0
## 26   26               Joakim Noah, CHI 11.7
## 27   27             Zach Randolph, MEM 15.7
## 28   28                  Rudy Gay, TOR 18.0
## 29   29              Kemba Walker, CHA 17.7
## 30   30          Chandler Parsons, HOU 14.6
## 31   31               Greg Monroe, DET 15.7
## 32   32                David West, IND 17.2
## 33   33               Monta Ellis, MIL 18.2
## 34   34                 O.J. Mayo, DAL 17.5
## 35   35                Marc Gasol, MEM 13.6
## 36   36                 Ty Lawson, DEN 16.3
## 37   37                Chris Paul, LAC 16.2
## 38   38            Greivis Vasquez, NO 13.6
## 39   39                Chris Bosh, MIA 17.1
## 40   40                Tim Duncan, SA 16.7
## 41   41               Joe Johnson, BKN 17.0
## 42   42          DeMarcus Cousins, SAC 17.3
## 43   43             DeMar DeRozan, TOR 17.5
## 44   44               Evan Turner, PHI 14.0
## 45   45          Danilo Gallinari, DEN 16.9
## 46   46              Klay Thompson, GS 16.3
## 47   47              Paul Millsap, UTA 15.1
## 48   48            Nikola Vucevic, ORL 12.3
## 49   49               Brook Lopez, BKN 18.7
## 50   50                George Hill, IND 14.6
```

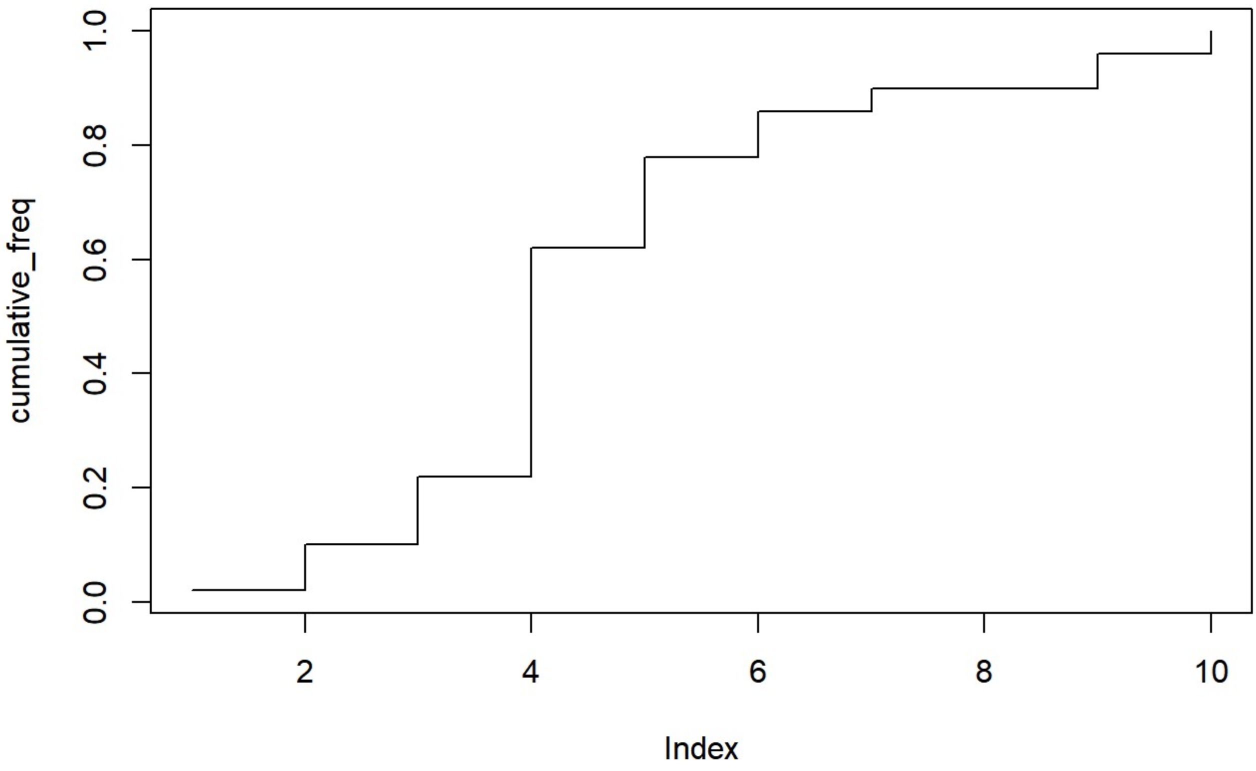
```
# a. Show the frequency distribution.
breaks <- seq(10, 30, by = 2)
hist(nba_players$PPG, breaks = breaks)
```



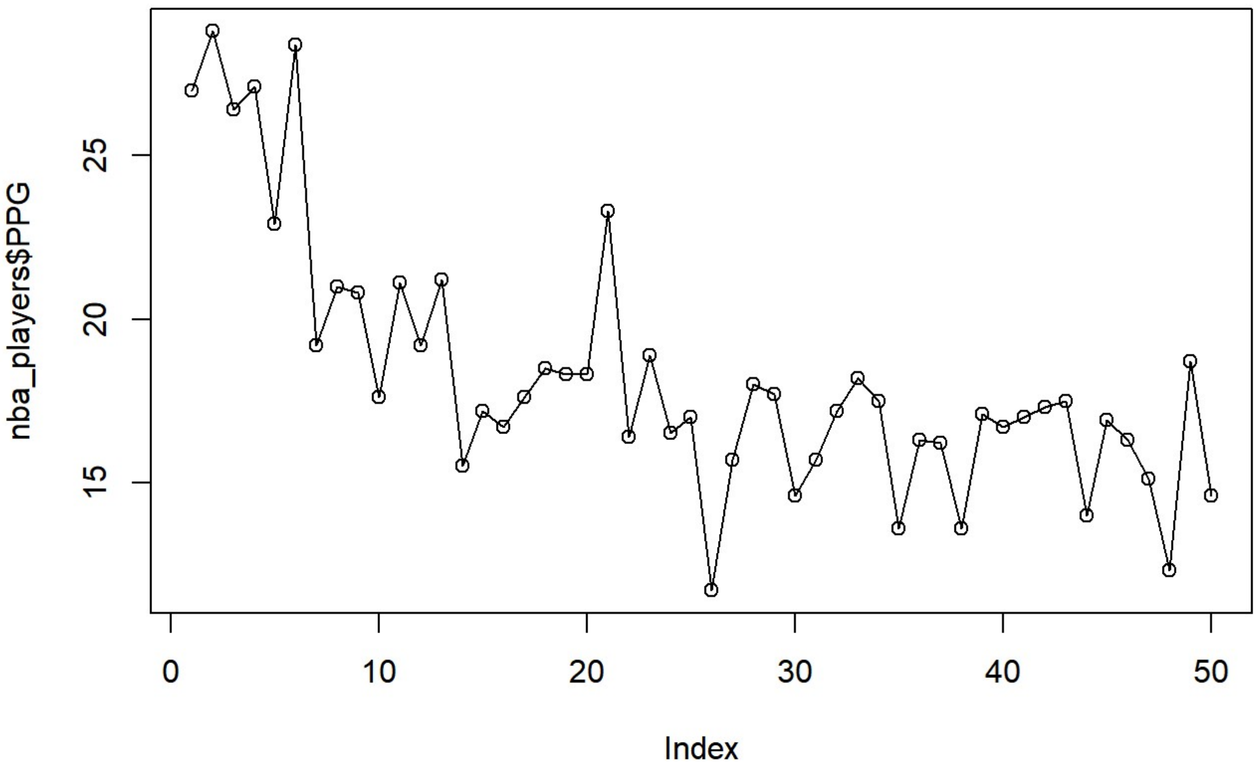

```
# b. Show the relative frequency distribution.
hist(nba_players$PPG, breaks = breaks, freq = FALSE)
```



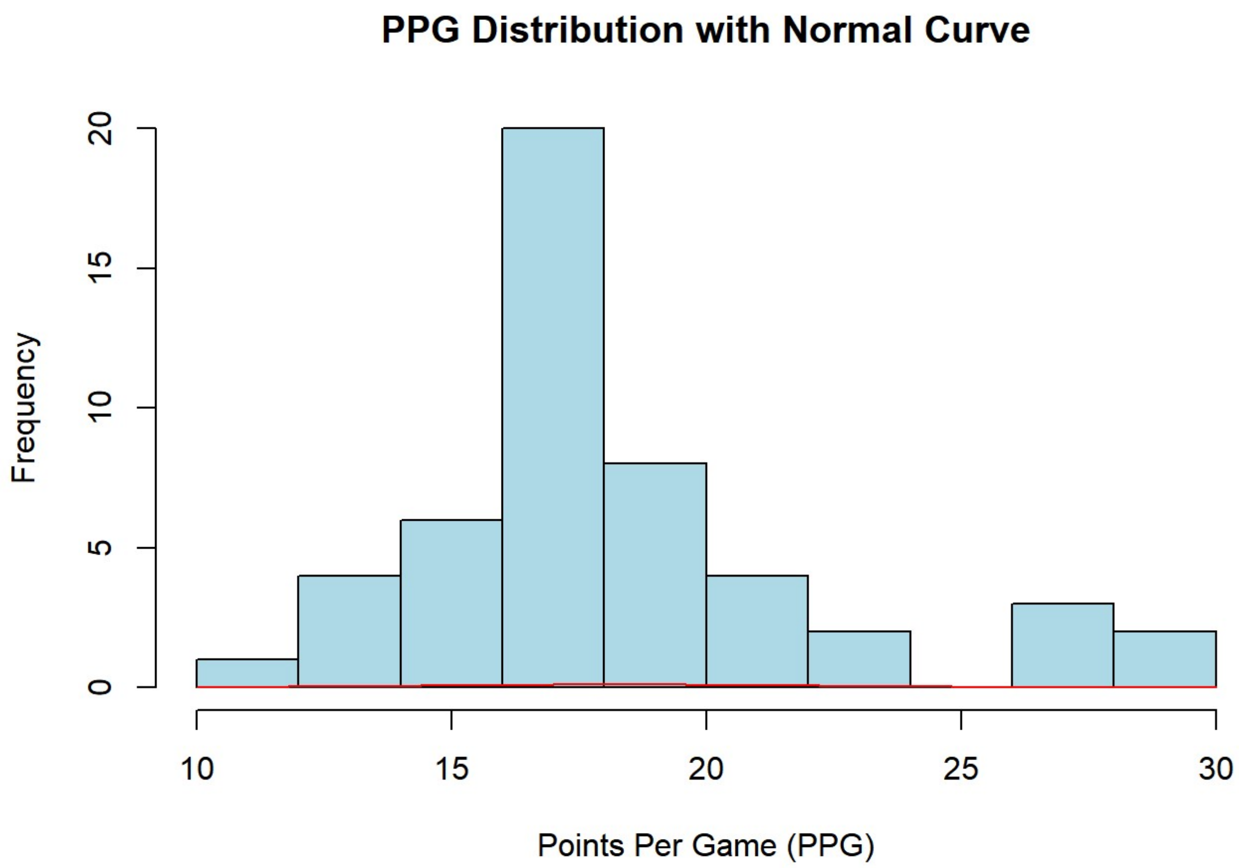

```
# c. Show the cumulative percent frequency distribution.
cumulative_freq <- cumsum(hist(nba_players$PPG, breaks = breaks, plot = FALSE)$counts) / s
um(hist(nba_players$PPG, breaks = breaks, plot = FALSE)$counts)
plot(cumulative_freq, type = "s")
```




```
# d. Develop a histogram for the average number of points scored per game
plot(nba_players$PPG, type="o")
```




```
# e. Do the data appear to be skewed? Explain.
hist(nba_players$PPG, main="PPG Distribution with Normal Curve", xlab="Points Per Game (PP
G)", col="lightblue")
curve(dnorm(x, mean=mean(nba_players$PPG), sd=sd(nba_players$PPG)), add=TRUE, col="red")
```




```
# f. What percentage of the players averaged at least 20 points per game?
count_players <- sum(nba_players$PPG >= 20)
count_players
```

```
## [1] 11
```

Question #3: A researcher reports survey results by stating that the standard error of the mean is 20. The population standard deviation is 500.

```
# a. How large was the sample used in this survey?
sigma <- 500
SE <- 20
n <- (sigma / SE) ^ 2
n
```

```
## [1] 625
```

```
# b. What is the probability that the point estimate was within ±25 of the population mea
n?
mu <- 0
x <- 25
z <- (x - mu) / SE
probability <- pnorm(z) - pnorm(-z)
probability
```

```
## [1] 0.7887005
```

Question #4: Young Professional Magazine (Attached Data: Professional)

Young Professional magazine was developed for a target audience of recent college graduates who are in their first 10 years in a business/professional career. In its two years of publication, the magazine has been fairly successful. Now the publisher is interested in expanding the magazine's advertising base. Potential advertisers continually ask about the demographics and interests of subscribers to young Professionals. To collect this information, the magazine commissioned a survey to develop a profile of its subscribers. The survey results will be used to help the magazine choose articles of interest and provide advertisers with a profile of subscribers. As a new employee of the magazine, you have been asked to help analyze the survey results. Some of the survey questions follow: 1. What is your age? 2. Are you: Male_________ Female___________ 3. Do you plan to make any real estate purchases in the next two years? Yes______ No______ 4. What is the approximate total value of financial investments, exclusive of your home, owned by you or members of your household? 5. How many stock/bond/mutual fund transactions have you made in the past year? 6. Do you have broadband access to the Internet at home? Yes______ No______ 7. Please indicate your total household income last year. ***8. Do you have children? Yes***___ No______ The file entitled Professional contains the responses to these questions. Managerial Report: Prepare a managerial report summarizing the results of the survey. In addition to statistical summaries, discuss how the magazine might use these results to attract advertisers. You might also comment on how the survey results could be used by the magazine's editors to identify topics that would be of interest to readers. Your report should address the following issues, but do not limit your analysis to just these areas.

```
data <- read.csv("data/Professional.csv", header = TRUE, sep = ",")

#a. Develop appropriate descriptive statistics to summarize the data.
colnames(data) <- c("age", "gender", "real_estate","investments","num_trans","has_broadban
d", "income", "have_children")
Professional <- data
Professional
```

```
##    age gender real_estate investments num_trans has_broadband income
## 1   38 Female          No       12200         4           Yes  75200
## 2   30   Male          No       12400         4           Yes  70300
## 3   41 Female          No       26800         5           Yes  48200
## 4   28 Female         Yes       19600         6            No  95300
## 5   31 Female         Yes       15100         5            No  73300
## 6   32   Male          No       39700         3           Yes 123400
## 7   32   Male         Yes       21900         2           Yes  73900
## 8   26 Female         Yes       41900         2           Yes  54300
## 9   26   Male         Yes       16100         4           Yes  93100
## 10  34 Female         Yes       18400        11           Yes  60100
## 11  33 Female          No       33800         3            No  48600
## 12  35 Female         Yes       15500         6            No  43500
## 13  28 Female          No       17300         7           Yes  73600
## 14  30   Male          No       47900         5           Yes  68200
## 15  30 Female          No       28200         3            No  61900
## 16  30   Male         Yes       19400         6            No  57600
## 17  33   Male         Yes       31000        12            No  82300
## 18  28   Male          No       21300         6            No  64600
## 19  27 Female          No       21300        10           Yes  61100
## 20  23 Female          No       21300         7           Yes  31200
## 21  30 Female          No       34100         6           Yes  92600
## 22  28   Male          No       32600         7           Yes  68300
## 23  41   Male         Yes           0        10           Yes  35100
## 24  29   Male         Yes       20800        10           Yes  85700
## 25  33 Female         Yes       23100         6           Yes 140300
## 26  30   Male          No       39800         9           Yes 108200
## 27  29 Female          No       17900         2            No  61100
## 28  33   Male          No       33300         2            No  33900
## 29  30 Female         Yes       21800         8           Yes  54400
## 30  30   Male          No       54000         7           Yes  61200
## 31  36 Female         Yes       34800        12            No  58000
## 32  33 Female         Yes       36100         7           Yes  90700
## 33  28 Female         Yes       44300         5           Yes  95200
## 34  28   Male          No       21400         5            No  50500
## 35  28   Male         Yes        8600         6           Yes  33800
## 36  35   Male          No       23200         3            No 147400
## 37  31   Male          No       24800         4            No  92600
## 38  33 Female          No       26600         5           Yes  66200
## 39  32   Male          No       33100         8            No  45700
## 40  28   Male          No       27000         4            No  60500
## 41  27   Male          No       48700         6            No 110600
## 42  30 Female         Yes       17100         4            No  60300
## 43  29   Male          No       19900         6            No  75700
## 44  28   Male          No       13200         9            No  70100
## 45  29 Female         Yes       32400         6            No  42100
## 46  32 Female          No       14200         5           Yes  41700
## 47  33   Male          No       20100         5            No  96900
## 48  23   Male         Yes       32000         5           Yes  65700
## 49  29 Female          No       41900         3            No  50200
## 50  31   Male          No       12000         3            No  61700
## 51  33 Female          No       14000         6            No  44500
## 52  37 Female          No       10000         4           Yes  51900
## 53  28 Female          No       27200         5           Yes 119100
```

```
## 54  27 Female          Yes       34500         4          Yes  49200
## 55  33 Female           No       26000         6          Yes  39000
## 56  31 Female           No       53400         3          Yes  35000
## 57  30   Male           No       23100         7          Yes 104700
## 58  23   Male           No       24900         5           No  49300
## 59  33   Male           No       10900        10          Yes  74000
## 60  31 Female           No       52500         2          Yes  57100
## 61  24 Female           No       24600        16          Yes  51400
## 62  35 Female           No       28800         6           No  62100
## 63  36 Female          Yes       37300        11          Yes 103000
## 64  24   Male          Yes       21100         9          Yes  97900
## 65  25   Male           No       40900         3           No 123100
## 66  26 Female           No       24700         7          Yes 322500
## 67  31   Male           No       33600         8          Yes  54800
## 68  26   Male          Yes       20000         6           No  66500
## 69  26   Male           No       23900         2           No  33700
## 70  28   Male          Yes       11700         8           No  73600
## 71  26 Female           No       19200         5          Yes  71300
## 72  34   Male           No       11900         6           No  74200
## 73  32   Male           No       17700         8          Yes  70000
## 74  30   Male           No       23900         6          Yes  40800
## 75  29 Female          Yes       27500         4          Yes  72500
## 76  23   Male          Yes       25300         4           No  53300
## 77  28 Female           No       66900         5           No  45600
## 78  25   Male           No       18800         5          Yes  73900
## 79  35   Male           No       24400         3          Yes  83600
## 80  31   Male           No       45500         9          Yes 124700
## 81  35   Male          Yes       54700         3          Yes 101600
## 82  32 Female           No       32200         4           No 205900
## 83  29   Male           No       16200         3          Yes  69700
## 84  34 Female           No       16000         9           No  95700
## 85  30 Female          Yes       24000         3           No  46100
## 86  29   Male           No       28800         3          Yes 118600
## 87  42   Male           No       14100         5          Yes  65400
## 88  37 Female          Yes       17700        11          Yes 149300
## 89  32   Male           No       29800         7          Yes 125000
## 90  33 Female          Yes       27200         3          Yes  39800
## 91  36   Male          Yes       43000         2           No  83500
## 92  32 Female          Yes       15500         7          Yes  38700
## 93  35   Male          Yes        8700         6           No 102400
## 94  20   Male           No       13900        14          Yes  57700
## 95  30   Male           No       14200         6          Yes  16200
## 96  31 Female          Yes       27100         3           No  43100
## 97  28   Male          Yes       22300         3           No  43700
## 98  24 Female          Yes       29500         4          Yes  39600
## 99  30   Male           No       18800         2          Yes 127500
## 100 23 Female          Yes       26400         9          Yes  33500
## 101 29   Male          Yes       16400         5          Yes  48100
## 102 33   Male           No       39500         3          Yes  52800
## 103 24   Male          Yes       20900         6          Yes  54800
## 104 25 Female           No       52800         5          Yes  46500
## 105 38   Male           No       36500         4          Yes  60400
## 106 22 Female           No       30400        11          Yes 202400
## 107 32 Female           No       39800         7           No  71300
## 108 28 Female          Yes       18100         9           No  62800
```

```
## 109  32   Male         Yes       69500         7           No  43900
## 110  33 Female          No       33000         5          Yes  52200
## 111  32 Female         Yes       32100         3           No  64400
## 112  24   Male         Yes       15200         8           No  77400
## 113  23 Female         Yes       49800         3          Yes  65900
## 114  23   Male          No       28100         6           No  54100
## 115  38   Male          No       15200         6          Yes  77400
## 116  29   Male          No       30800         3          Yes  85900
## 117  28   Male          No       21200         3           No 148600
## 118  29   Male          No       24600         4          Yes  82100
## 119  36   Male         Yes       24700         9          Yes  64400
## 120  35 Female         Yes       26300        16           No  86200
## 121  34 Female         Yes       32200         6           No 177100
## 122  23   Male          No       20300         6          Yes  68300
## 123  31 Female          No       60900         6          Yes  67900
## 124  28   Male          No       15800        11          Yes  57300
## 125  33 Female          No       31700         6          Yes  83600
## 126  31   Male         Yes           0         5          Yes  77500
## 127  26 Female         Yes       20900         9          Yes  61700
## 128  38 Female          No       37900        12           No  85600
## 129  28   Male         Yes       22400         7          Yes  59900
## 130  27   Male          No       26600         5          Yes  43100
## 131  31   Male         Yes       20800        10           No  65700
## 132  34   Male          No       31100         7          Yes  80500
## 133  29   Male          No       31200         5          Yes  88400
## 134  27   Male          No           0         2           No  40200
## 135  26   Male         Yes       38900         7          Yes  84000
## 136  33   Male         Yes       29500         4          Yes  34400
## 137  33   Male          No       16100         5           No  55800
## 138  34   Male          No       15800         7          Yes  64900
## 139  31   Male          No       27500         4          Yes  41100
## 140  37 Female          No       13000         7           No  39100
## 141  27   Male         Yes       68100         5           No  58200
## 142  32   Male         Yes       18500         7          Yes  46500
## 143  25 Female          No       27100         2          Yes  45100
## 144  31 Female          No       25900         5          Yes  69900
## 145  38   Male          No       18400         2           No  63700
## 146  31 Female          No       59200         4          Yes  40500
## 147  28 Female          No       67900         5           No  62600
## 148  29 Female          No       36700         7          Yes  73400
## 149  34 Female          No       32400         9          Yes  35000
## 150  35 Female         Yes       10300         4          Yes 114000
## 151  30   Male          No       39300         7          Yes  70800
## 152  32 Female          No       13800         6          Yes  60300
## 153  31   Male          No       11100         4          Yes  78700
## 154  29 Female          No       28600         2           No 179700
## 155  31   Male          No       22100         6          Yes 157200
## 156  31 Female          No       16700         6          Yes  67300
## 157  22 Female         Yes       19400         6          Yes  49800
## 158  19   Male         Yes       54600         4          Yes  43400
## 159  31 Female         Yes       41500         2          Yes  71100
## 160  22 Female          No       23300         9           No  42100
## 161  26   Male          No       29700         8          Yes  42200
## 162  34   Male         Yes       33200         2          Yes  63300
## 163  37   Male         Yes       56300         5           No  55700
```

```
## 164  36 Female          No       25900         5          Yes 107900
## 165  27 Female         Yes       63500         6          Yes  44800
## 166  32   Male         Yes       18700         7          Yes  94700
## 167  27 Female          No       49700         4          Yes 112700
## 168  32   Male         Yes       15000         5          Yes 114900
## 169  24 Female         Yes        8200         7           No 112700
## 170  38 Female         Yes       36300         9           No  60900
## 171  33 Female          No       50200         7          Yes 137800
## 172  19   Male          No       13500        10          Yes  47200
## 173  33   Male         Yes       37500         4          Yes  70900
## 174  25 Female          No       26800         7           No  52000
## 175  27   Male          No       30300         5          Yes 166500
## 176  35 Female          No       46300        18          Yes  88000
## 177  29   Male         Yes       14400         9           No  41800
## 178  31   Male         Yes           0         4          Yes  60300
## 179  29 Female         Yes       18800         5          Yes  53400
## 180  31 Female          No       40900         5           No  46700
## 181  32   Male         Yes       23300         7          Yes  59600
## 182  22 Female          No       11900         4          Yes  27700
## 183  29   Male         Yes       28100        10           No  48900
## 184  28 Female          No       20400         8           No  56600
## 185  26   Male          No       14900         7           No  32500
## 186  33 Female          No       33900         3          Yes  74500
## 187  35   Male          No       15100         6           No  69800
## 188  27   Male          No      133400         5          Yes  48100
## 189  33 Female         Yes       18300         7          Yes  95900
## 190  30 Female         Yes       28700        10          Yes  76500
## 191  28   Male         Yes       50900         6          Yes  43700
## 192  35 Female          No       25600         4          Yes  72700
## 193  34 Female         Yes       30700         0          Yes 140100
## 194  28   Male         Yes       64300         5          Yes  76600
## 195  30   Male         Yes       67200         7           No  65300
## 196  37   Male          No       62200        11          Yes 161100
## 197  32   Male          No       33100        12           No  64900
## 198  21   Male          No       26000         4           No  59800
## 199  31 Female         Yes       18400         5           No  87900
## 200  31   Male         Yes       53000        12          Yes 174900
## 201  23   Male          No       46300         3          Yes  40700
## 202  34   Male         Yes       32700         5           No  36700
## 203  31   Male          No       50900         4          Yes 167800
## 204  28 Female          No       19200         2          Yes 168500
## 205  30   Male         Yes       18300         3          Yes 112100
## 206  25   Male          No       17800         2          Yes  65500
## 207  32 Female         Yes       41800         6          Yes  73500
## 208  30   Male         Yes        7100         7          Yes  59100
## 209  28   Male         Yes       15500         8          Yes  73900
## 210  33   Male          No       22900         4          Yes 120800
## 211  34 Female         Yes       10000         4           No  25100
## 212  34 Female          No       53300         6          Yes 137900
## 213  33 Female         Yes       32400         7          Yes  70200
## 214  26   Male          No       85600         2          Yes  90900
## 215  28 Female         Yes       26400         6           No  65500
## 216  23   Male         Yes       30400         5          Yes  86200
## 217  31   Male         Yes       24000         8          Yes  45900
## 218  28 Female          No       17500         0           No  76000
```

```
## 219  29   Male          No       11000          5           Yes  82700
## 220  30 Female          No        9400          5            No  57400
## 221  30   Male          No       34700         12            No  53700
## 222  31 Female         Yes       44200          3            No  29900
## 223  22   Male          No       20700          5            No  94800
## 224  34   Male          No       29000          3            No  67300
## 225  30 Female          No       29800          3           Yes  41200
## 226  34 Female         Yes       95200         21            No  70400
## 227  25   Male         Yes       21500          5           Yes 120600
## 228  32 Female          No       23100          4            No  63900
## 229  31 Female         Yes       13900          5           Yes  98800
## 230  29 Female          No       28900          7           Yes  53300
## 231  28 Female          No       70400          8            No  30000
## 232  26 Female          No       12400          4           Yes  93200
## 233  26   Male         Yes       27800          9           Yes  84100
## 234  26 Female          No       19900          5           Yes 124000
## 235  29 Female          No       28000          7           Yes  73400
## 236  32 Female         Yes       75000          6           Yes  45500
## 237  37 Female          No       14000          9           Yes  88400
## 238  32   Male          No       20900          4           Yes  74900
## 239  33 Female          No       47200          6           Yes  51700
## 240  36 Female          No       57200          5            No  78700
## 241  30   Male         Yes       42900          7           Yes  30700
## 242  31   Male         Yes       44200          1           Yes  56000
## 243  28 Female         Yes       28000          4           Yes  88400
## 244  25   Male         Yes       24500          6           Yes 201700
## 245  29   Male          No       18100          9           Yes  71600
## 246  31 Female          No       20900          7            No  59000
## 247  32   Male          No       32100          6            No  57300
## 248  32   Male         Yes       44500          6            No  91900
## 249  38   Male         Yes       20100          4           Yes  68300
## 250  35   Male         Yes       36000          8            No  93900
## 251  34   Male          No       12600         12           Yes  61200
## 252  33   Male         Yes       19300          6            No 128800
## 253  32 Female         Yes       33300         13           Yes  77900
## 254  26 Female          No       23600          2            No  67700
## 255  23 Female          No       49400          7           Yes  63300
## 256  32   Male          No       23700          1           Yes  55000
## 257  25 Female         Yes       35700          4            No 100700
## 258  27   Male         Yes       19900         10            No 148000
## 259  22   Male          No       30600          9            No  59600
## 260  33 Female          No       24600          2            No  51600
## 261  33   Male         Yes       18000          1           Yes  48400
## 262  29   Male          No           0          6           Yes  58500
## 263  33 Female          No       27600          4           Yes 153000
## 264  33 Female          No       48800          4            No  86800
## 265  42   Male          No       30300          4            No  46400
## 266  28 Female         Yes       57700          6           Yes  36900
## 267  27 Female          No       29200          3           Yes  46900
## 268  35   Male         Yes       16600          2           Yes  57400
## 269  31 Female          No       16800          5            No  91200
## 270  32 Female         Yes       21400          8           Yes  62100
## 271  34   Male          No       12500          7           Yes  38800
## 272  24 Female         Yes       49300         13            No  90200
## 273  30   Male          No        9100          5           Yes  47100
```

```
## 274  28   Male          No       19200          8            No 159200
## 275  32 Female         Yes       15900          5           Yes  52700
## 276  34   Male         Yes       40200         13           Yes  73400
## 277  31   Male         Yes       55800          5           Yes  58900
## 278  33   Male         Yes       27200          3           Yes  81100
## 279  31 Female         Yes       20100          4            No  68400
## 280  29   Male          No       17300          3            No  55700
## 281  25   Male         Yes       42100          7            No  45400
## 282  29   Male          No       39500          5           Yes  59600
## 283  25 Female          No       21400          2           Yes  47700
## 284  30 Female          No       10200         11           Yes  57200
## 285  28 Female         Yes       18300         11           Yes  47300
## 286  32 Female          No       20600          4           Yes  58300
## 287  34 Female          No       63400         17           Yes  47800
## 288  33 Female         Yes       38200          5           Yes  61600
## 289  24   Male          No       50400          7           Yes  64000
## 290  30 Female          No        9500          5            No  87400
## 291  25   Male          No       27900          6           Yes  56200
## 292  26   Male          No       18100          4           Yes 100600
## 293  31   Male          No       39500          5           Yes  51100
## 294  29   Male          No       23400          5           Yes  90100
## 295  31   Male          No       34100          6           Yes  50800
## 296  36   Male          No       17500          6            No  62500
## 297  34 Female         Yes       12600          7            No  37400
## 298  24   Male          No       29200          3           Yes 102700
## 299  28   Male         Yes       42600          4            No  68400
## 300  34 Female         Yes       18100          4            No  68800
## 301  31   Male          No       11200          7           Yes 108000
## 302  31   Male         Yes       30500          9           Yes 200500
## 303  30   Male          No       39300          9            No 123000
## 304  38 Female          No       23100          7           Yes  64300
## 305  29   Male          No       38400          2           Yes  72600
## 306  21   Male         Yes       35300          6           Yes  81500
## 307  32 Female          No       15800          3            No  57000
## 308  27   Male         Yes       24300          6           Yes 101500
## 309  29   Male         Yes       24700          2           Yes  67600
## 310  32   Male          No       30800          7            No  70200
## 311  32   Male         Yes       17400          3            No  64800
## 312  29 Female          No        9100          2           Yes  70800
## 313  35   Male         Yes       23300          2           Yes  50300
## 314  36   Male         Yes       15000          2           Yes  30300
## 315  32   Male          No           0         14            No  61700
## 316  25 Female         Yes       31300          7            No 103500
## 317  29   Male         Yes       34500          8            No  33400
## 318  22   Male         Yes       26300          7           Yes  51700
## 319  28   Male          No       21900          7           Yes  76400
## 320  29 Female         Yes       27100          7           Yes  47400
## 321  29   Male          No       59300         11           Yes 118800
## 322  29 Female          No       34500          9           Yes  61300
## 323  32 Female         Yes       63200          8            No  97900
## 324  30   Male          No        7000          3           Yes  67400
## 325  29 Female          No       18100          7            No  82000
## 326  30 Female          No       15500          5           Yes  60700
## 327  38   Male         Yes       11800          4            No  98800
## 328  32 Female         Yes       27700          3           Yes  61800
```

```
## 329  27   Male          No        14600         6          No  49600
## 330  29   Male          No        17500         3         Yes  54700
## 331  28 Female         Yes        37900         7         Yes 153300
## 332  33   Male         Yes        13800         9         Yes  94400
## 333  32   Male         Yes        22100         9         Yes  39500
## 334  35 Female         Yes        24200         6         Yes  72400
## 335  27   Male         Yes        19500         6         Yes  93200
## 336  29   Male         Yes        23300         4         Yes 161400
## 337  30   Male          No        29000         5          No  56100
## 338  34 Female          No        53200         5          No  95000
## 339  24   Male          No        15300         6         Yes  94100
## 340  30 Female         Yes        25800         8          No  24300
## 341  28   Male          No        34200         4         Yes  49400
## 342  23 Female          No        25000         7         Yes  41200
## 343  26 Female         Yes        24500         7          No  81800
## 344  30   Male          No        25500         6          No  62400
## 345  26   Male          No        19400        14          No 117800
## 346  24 Female          No        30600        11         Yes 100100
## 347  33 Female          No        21100         9          No  94100
## 348  30 Female          No        83300        13         Yes  59300
## 349  25 Female         Yes        50700         5         Yes  89500
## 350  24   Male          No        29600         2          No  88900
## 351  35   Male         Yes        29700         6         Yes  53200
## 352  34 Female         Yes        12100         6          No  58600
## 353  36   Male          No        25000         2          No  54200
## 354  29   Male          No        37200         5         Yes  38500
## 355  27 Female          No        16600         1          No  69800
## 356  30   Male          No        58000         2         Yes  66200
## 357  25   Male         Yes        23400         4         Yes  78900
## 358  32   Male         Yes        21500         3          No  87200
## 359  29   Male         Yes        33800        15         Yes  79300
## 360  27   Male          No        18400         7          No  51800
## 361  27   Male         Yes        13100        11         Yes  70900
## 362  33   Male          No        18100        18         Yes 110900
## 363  29   Male         Yes        34300         2          No 109400
## 364  30   Male         Yes        23500         8          No  43800
## 365  32   Male         Yes        19200         6          No  49700
## 366  23 Female         Yes        16900         4         Yes 149300
## 367  26   Male          No        22100         5         Yes  68900
## 368  34 Female         Yes        27000         3         Yes  50100
## 369  35   Male         Yes        65200         2          No  32400
## 370  31 Female         Yes        11200         5         Yes  82700
## 371  29 Female          No        24800        12          No 104300
## 372  25   Male         Yes        46500         9         Yes  49600
## 373  34   Male         Yes        20000         7         Yes  36500
## 374  38 Female          No        16200         1         Yes  65500
## 375  34   Male          No        27200         9         Yes 110800
## 376  32 Female          No        11100         3          No  38100
## 377  33 Female         Yes        18500         2         Yes  70700
## 378  27   Male          No        21300         1         Yes  39800
## 379  30   Male          No        19400         3         Yes  60600
## 380  32 Female         Yes        17000         5          No  48200
## 381  33   Male          No        28700         4          No 108700
## 382  26 Female         Yes        20700         4          No  71300
## 383  29   Male          No        17100         4         Yes 110800
```

```
## 384  36   Male          No       21000          5           Yes 120300
## 385  38   Male         Yes       21900          6           Yes 130000
## 386  31 Female         Yes       61200          7           Yes  60300
## 387  31 Female          No       10200          8            No  70300
## 388  31   Male         Yes       23600          5            No 100700
## 389  30   Male          No       23900          4            No  72300
## 390  30   Male          No       28100          5           Yes  45800
## 391  24   Male          No       23600          6            No  51600
## 392  34   Male          No       70800         10           Yes  48200
## 393  39   Male          No       25100         14            No  87900
## 394  27   Male          No       22300          7            No  89600
## 395  31 Female         Yes       36200          6           Yes  56100
## 396  32   Male         Yes       20600         12           Yes  93600
## 397  37   Male         Yes       38300          4            No 100300
## 398  29   Male          No       13800          5            No  96000
## 399  31   Male         Yes       47000         10           Yes 105500
## 400  27   Male          No       30300          4            No  76400
## 401  31 Female         Yes       37100          6           Yes  77100
## 402  32 Female         Yes       64100          4           Yes  62300
## 403  30   Male         Yes       19300          6            No  52700
## 404  26 Female         Yes       27400          4           Yes  59000
## 405  32 Female         Yes       23900          2           Yes  42200
## 406  32 Female          No       27400          7           Yes  83500
## 407  26 Female         Yes       14400          3           Yes  28200
## 408  24   Male         Yes       36000          4           Yes 103500
## 409  25 Female         Yes       30300          6            No  35100
## 410  28 Female         Yes       30500         11           Yes  36800
##     have_children NA NA NA NA NA NA
## 1             Yes NA    NA NA NA NA
## 2             Yes NA    NA NA NA NA
## 3              No NA    NA NA NA NA
## 4              No NA    NA NA NA NA
## 5             Yes NA  . NA NA NA NA
## 6             Yes NA    NA NA NA NA
## 7             Yes NA    NA NA NA NA
## 8             Yes NA    NA NA NA NA
## 9              No NA    NA NA NA NA
## 10             No NA    NA NA NA NA
## 11             No NA    NA NA NA NA
## 12             No NA    NA NA NA NA
## 13            Yes NA    NA NA NA NA
## 14            Yes NA    NA NA NA NA
## 15            Yes NA    NA NA NA NA
## 16             No NA    NA NA NA NA
## 17             No NA    NA NA NA NA
## 18            Yes NA    NA NA NA NA
## 19            Yes NA    NA NA NA NA
## 20            Yes NA    NA NA NA NA
## 21             No NA    NA NA NA NA
## 22             No NA    NA NA NA NA
## 23             No NA    NA NA NA NA
## 24             No NA    NA NA NA NA
## 25            Yes NA    NA NA NA NA
## 26            Yes NA    NA NA NA NA
## 27             No NA    NA NA NA NA
```

```
## 28            Yes NA    NA NA NA NA
## 29            Yes NA    NA NA NA NA
## 30             No NA    NA NA NA NA
## 31            Yes NA    NA NA NA NA
## 32            Yes NA    NA NA NA NA
## 33             No NA    NA NA NA NA
## 34            Yes NA    NA NA NA NA
## 35            Yes NA    NA NA NA NA
## 36             No NA    NA NA NA NA
## 37            Yes NA    NA NA NA NA
## 38             No NA    NA NA NA NA
## 39             No NA    NA NA NA NA
## 40            Yes NA    NA NA NA NA
## 41             No NA    NA NA NA NA
## 42            Yes NA    NA NA NA NA
## 43             No NA    NA NA NA NA
## 44            Yes NA    NA NA NA NA
## 45            Yes NA    NA NA NA NA
## 46            Yes NA    NA NA NA NA
## 47             No NA    NA NA NA NA
## 48             No NA    NA NA NA NA
## 49             No NA    NA NA NA NA
## 50             No NA    NA NA NA NA
## 51            Yes NA    NA NA NA NA
## 52             No NA    NA NA NA NA
## 53            Yes NA    NA NA NA NA
## 54             No NA    NA NA NA NA
## 55            Yes NA    NA NA NA NA
## 56             No NA    NA NA NA NA
## 57            Yes NA    NA NA NA NA
## 58             No NA    NA NA NA NA
## 59             No NA    NA NA NA NA
## 60            Yes NA    NA NA NA NA
## 61            Yes NA    NA NA NA NA
## 62            Yes NA    NA NA NA NA
## 63            Yes NA    NA NA NA NA
## 64             No NA    NA NA NA NA
## 65             No NA    NA NA NA NA
## 66             No NA    NA NA NA NA
## 67            Yes NA    NA NA NA NA
## 68            Yes NA    NA NA NA NA
## 69            Yes NA    NA NA NA NA
## 70            Yes NA    NA NA NA NA
## 71            Yes NA    NA NA NA NA
## 72             No NA    NA NA NA NA
## 73            Yes NA    NA NA NA NA
## 74             No NA    NA NA NA NA
## 75            Yes NA    NA NA NA NA
## 76             No NA    NA NA NA NA
## 77             No NA    NA NA NA NA
## 78            Yes NA    NA NA NA NA
## 79             No NA    NA NA NA NA
## 80             No NA    NA NA NA NA
## 81             No NA    NA NA NA NA
## 82            Yes NA    NA NA NA NA
```

```
## 83             No NA    NA NA NA NA
## 84             No NA    NA NA NA NA
## 85            Yes NA    NA NA NA NA
## 86             No NA    NA NA NA NA
## 87             No NA    NA NA NA NA
## 88            Yes NA    NA NA NA NA
## 89             No NA    NA NA NA NA
## 90             No NA    NA NA NA NA
## 91            Yes NA    NA NA NA NA
## 92            Yes NA    NA NA NA NA
## 93             No NA    NA NA NA NA
## 94             No NA    NA NA NA NA
## 95             No NA    NA NA NA NA
## 96             No NA    NA NA NA NA
## 97             No NA    NA NA NA NA
## 98            Yes NA    NA NA NA NA
## 99             No NA    NA NA NA NA
## 100            No NA    NA NA NA NA
## 101           Yes NA    NA NA NA NA
## 102            No NA    NA NA NA NA
## 103            No NA    NA NA NA NA
## 104            No NA    NA NA NA NA
## 105            No NA    NA NA NA NA
## 106           Yes NA    NA NA NA NA
## 107            No NA    NA NA NA NA
## 108            No NA    NA NA NA NA
## 109           Yes NA    NA NA NA NA
## 110           Yes NA    NA NA NA NA
## 111           Yes NA    NA NA NA NA
## 112           Yes NA    NA NA NA NA
## 113            No NA    NA NA NA NA
## 114           Yes NA    NA NA NA NA
## 115           Yes NA    NA NA NA NA
## 116           Yes NA    NA NA NA NA
## 117            No NA    NA NA NA NA
## 118            No NA    NA NA NA NA
## 119           Yes NA    NA NA NA NA
## 120            No NA    NA NA NA NA
## 121           Yes NA    NA NA NA NA
## 122            No NA    NA NA NA NA
## 123           Yes NA    NA NA NA NA
## 124           Yes NA    NA NA NA NA
## 125            No NA    NA NA NA NA
## 126            No NA    NA NA NA NA
## 127           Yes NA    NA NA NA NA
## 128           Yes NA    NA NA NA NA
## 129           Yes NA    NA NA NA NA
## 130            No NA    NA NA NA NA
## 131            No NA    NA NA NA NA
## 132           Yes NA    NA NA NA NA
## 133           Yes NA    NA NA NA NA
## 134            No NA    NA NA NA NA
## 135           Yes NA    NA NA NA NA
## 136           Yes NA    NA NA NA NA
## 137           Yes NA    NA NA NA NA
```

```
## 138            Yes NA    NA NA NA NA
## 139            Yes NA    NA NA NA NA
## 140            Yes NA    NA NA NA NA
## 141             No NA    NA NA NA NA
## 142            Yes NA    NA NA NA NA
## 143            Yes NA    NA NA NA NA
## 144            Yes NA    NA NA NA NA
## 145             No NA    NA NA NA NA
## 146             No NA    NA NA NA NA
## 147            Yes NA    NA NA NA NA
## 148            Yes NA    NA NA NA NA
## 149             No NA    NA NA NA NA
## 150             No NA    NA NA NA NA
## 151             No NA    NA NA NA NA
## 152            Yes NA    NA NA NA NA
## 153             No NA    NA NA NA NA
## 154             No NA    NA NA NA NA
## 155            Yes NA    NA NA NA NA
## 156             No NA    NA NA NA NA
## 157             No NA    NA NA NA NA
## 158            Yes NA    NA NA NA NA
## 159            Yes NA    NA NA NA NA
## 160            Yes NA    NA NA NA NA
## 161            Yes NA    NA NA NA NA
## 162             No NA    NA NA NA NA
## 163             No NA    NA NA NA NA
## 164             No NA    NA NA NA NA
## 165            Yes NA    NA NA NA NA
## 166             No NA    NA NA NA NA
## 167             No NA    NA NA NA NA
## 168            Yes NA    NA NA NA NA
## 169            Yes NA    NA NA NA NA
## 170             No NA    NA NA NA NA
## 171             No NA    NA NA NA NA
## 172            Yes NA    NA NA NA NA
## 173            Yes NA    NA NA NA NA
## 174             No NA    NA NA NA NA
## 175            Yes NA    NA NA NA NA
## 176             No NA    NA NA NA NA
## 177             No NA    NA NA NA NA
## 178             No NA    NA NA NA NA
## 179            Yes NA    NA NA NA NA
## 180            Yes NA    NA NA NA NA
## 181            Yes NA    NA NA NA NA
## 182            Yes NA    NA NA NA NA
## 183            Yes NA    NA NA NA NA
## 184             No NA    NA NA NA NA
## 185             No NA    NA NA NA NA
## 186            Yes NA    NA NA NA NA
## 187             No NA    NA NA NA NA
## 188            Yes NA    NA NA NA NA
## 189            Yes NA    NA NA NA NA
## 190             No NA    NA NA NA NA
## 191             No NA    NA NA NA NA
## 192            Yes NA    NA NA NA NA
```

```
## 193             No NA    NA NA NA NA
## 194             No NA    NA NA NA NA
## 195            Yes NA    NA NA NA NA
## 196            Yes NA    NA NA NA NA
## 197             No NA    NA NA NA NA
## 198            Yes NA    NA NA NA NA
## 199            Yes NA    NA NA NA NA
## 200            Yes NA    NA NA NA NA
## 201            Yes NA    NA NA NA NA
## 202            Yes NA    NA NA NA NA
## 203             No NA    NA NA NA NA
## 204             No NA    NA NA NA NA
## 205            Yes NA    NA NA NA NA
## 206             No NA    NA NA NA NA
## 207             No NA    NA NA NA NA
## 208            Yes NA    NA NA NA NA
## 209             No NA    NA NA NA NA
## 210            Yes NA    NA NA NA NA
## 211            Yes NA    NA NA NA NA
## 212            Yes NA    NA NA NA NA
## 213            Yes NA    NA NA NA NA
## 214             No NA    NA NA NA NA
## 215             No NA    NA NA NA NA
## 216             No NA    NA NA NA NA
## 217             No NA    NA NA NA NA
## 218             No NA    NA NA NA NA
## 219            Yes NA    NA NA NA NA
## 220            Yes NA    NA NA NA NA
## 221             No NA    NA NA NA NA
## 222            Yes NA    NA NA NA NA
## 223            Yes NA    NA NA NA NA
## 224            Yes NA    NA NA NA NA
## 225            Yes NA    NA NA NA NA
## 226            Yes NA    NA NA NA NA
## 227            Yes NA    NA NA NA NA
## 228             No NA    NA NA NA NA
## 229            Yes NA    NA NA NA NA
## 230            Yes NA    NA NA NA NA
## 231            Yes NA    NA NA NA NA
## 232            Yes NA    NA NA NA NA
## 233             No NA    NA NA NA NA
## 234            Yes NA    NA NA NA NA
## 235            Yes NA    NA NA NA NA
## 236             No NA    NA NA NA NA
## 237            Yes NA    NA NA NA NA
## 238             No NA    NA NA NA NA
## 239            Yes NA    NA NA NA NA
## 240            Yes NA    NA NA NA NA
## 241            Yes NA    NA NA NA NA
## 242            Yes NA    NA NA NA NA
## 243            Yes NA    NA NA NA NA
## 244            Yes NA    NA NA NA NA
## 245            Yes NA    NA NA NA NA
## 246            Yes NA    NA NA NA NA
## 247            Yes NA    NA NA NA NA
```

```
## 248          Yes NA    NA NA NA NA
## 249          Yes NA    NA NA NA NA
## 250           No NA    NA NA NA NA
## 251           No NA    NA NA NA NA
## 252          Yes NA    NA NA NA NA
## 253           No NA    NA NA NA NA
## 254          Yes NA    NA NA NA NA
## 255          Yes NA    NA NA NA NA
## 256           No NA    NA NA NA NA
## 257          Yes NA    NA NA NA NA
## 258          Yes NA    NA NA NA NA
## 259           No NA    NA NA NA NA
## 260           No NA    NA NA NA NA
## 261           No NA    NA NA NA NA
## 262          Yes NA    NA NA NA NA
## 263           No NA    NA NA NA NA
## 264          Yes NA    NA NA NA NA
## 265          Yes NA    NA NA NA NA
## 266           No NA    NA NA NA NA
## 267          Yes NA    NA NA NA NA
## 268           No NA    NA NA NA NA
## 269           No NA    NA NA NA NA
## 270          Yes NA    NA NA NA NA
## 271          Yes NA    NA NA NA NA
## 272          Yes NA    NA NA NA NA
## 273          Yes NA    NA NA NA NA
## 274          Yes NA    NA NA NA NA
## 275          Yes NA    NA NA NA NA
## 276           No NA    NA NA NA NA
## 277           No NA    NA NA NA NA
## 278          Yes NA    NA NA NA NA
## 279          Yes NA    NA NA NA NA
## 280          Yes NA    NA NA NA NA
## 281           No NA    NA NA NA NA
## 282           No NA    NA NA NA NA
## 283          Yes NA    NA NA NA NA
## 284          Yes NA    NA NA NA NA
## 285          Yes NA    NA NA NA NA
## 286          Yes NA    NA NA NA NA
## 287          Yes NA    NA NA NA NA
## 288           No NA    NA NA NA NA
## 289          Yes NA    NA NA NA NA
## 290          Yes NA    NA NA NA NA
## 291          Yes NA    NA NA NA NA
## 292           No NA    NA NA NA NA
## 293           No NA    NA NA NA NA
## 294           No NA    NA NA NA NA
## 295          Yes NA    NA NA NA NA
## 296           No NA    NA NA NA NA
## 297           No NA    NA NA NA NA
## 298           No NA    NA NA NA NA
## 299           No NA    NA NA NA NA
## 300           No NA    NA NA NA NA
## 301           No NA    NA NA NA NA
## 302           No NA    NA NA NA NA
```

```
## 303            No NA    NA NA NA NA
## 304           Yes NA    NA NA NA NA
## 305            No NA    NA NA NA NA
## 306            No NA    NA NA NA NA
## 307           Yes NA    NA NA NA NA
## 308            No NA    NA NA NA NA
## 309            No NA    NA NA NA NA
## 310            No NA    NA NA NA NA
## 311            No NA    NA NA NA NA
## 312            No NA    NA NA NA NA
## 313           Yes NA    NA NA NA NA
## 314           Yes NA    NA NA NA NA
## 315           Yes NA    NA NA NA NA
## 316            No NA    NA NA NA NA
## 317            No NA    NA NA NA NA
## 318            No NA    NA NA NA NA
## 319           Yes NA    NA NA NA NA
## 320           Yes NA    NA NA NA NA
## 321           Yes NA    NA NA NA NA
## 322           Yes NA    NA NA NA NA
## 323            No NA    NA NA NA NA
## 324           Yes NA    NA NA NA NA
## 325            No NA    NA NA NA NA
## 326            No NA    NA NA NA NA
## 327           Yes NA    NA NA NA NA
## 328            No NA    NA NA NA NA
## 329           Yes NA    NA NA NA NA
## 330           Yes NA    NA NA NA NA
## 331            No NA    NA NA NA NA
## 332           Yes NA    NA NA NA NA
## 333           Yes NA    NA NA NA NA
## 334           Yes NA    NA NA NA NA
## 335           Yes NA    NA NA NA NA
## 336            No NA    NA NA NA NA
## 337            No NA    NA NA NA NA
## 338           Yes NA    NA NA NA NA
## 339           Yes NA    NA NA NA NA
## 340           Yes NA    NA NA NA NA
## 341            No NA    NA NA NA NA
## 342            No NA    NA NA NA NA
## 343            No NA    NA NA NA NA
## 344            No NA    NA NA NA NA
## 345           Yes NA    NA NA NA NA
## 346           Yes NA    NA NA NA NA
## 347           Yes NA    NA NA NA NA
## 348           Yes NA    NA NA NA NA
## 349           Yes NA    NA NA NA NA
## 350           Yes NA    NA NA NA NA
## 351            No NA    NA NA NA NA
## 352            No NA    NA NA NA NA
## 353            No NA    NA NA NA NA
## 354            No NA    NA NA NA NA
## 355           Yes NA    NA NA NA NA
## 356            No NA    NA NA NA NA
## 357            No NA    NA NA NA NA
```

```
## 358            No NA    NA NA NA NA
## 359            No NA    NA NA NA NA
## 360            No NA    NA NA NA NA
## 361            No NA    NA NA NA NA
## 362            No NA    NA NA NA NA
## 363           Yes NA    NA NA NA NA
## 364            No NA    NA NA NA NA
## 365           Yes NA    NA NA NA NA
## 366           Yes NA    NA NA NA NA
## 367            No NA    NA NA NA NA
## 368            No NA    NA NA NA NA
## 369           Yes NA    NA NA NA NA
## 370           Yes NA    NA NA NA NA
## 371           Yes NA    NA NA NA NA
## 372            No NA    NA NA NA NA
## 373            No NA    NA NA NA NA
## 374           Yes NA    NA NA NA NA
## 375           Yes NA    NA NA NA NA
## 376            No NA    NA NA NA NA
## 377           Yes NA    NA NA NA NA
## 378           Yes NA    NA NA NA NA
## 379           Yes NA    NA NA NA NA
## 380           Yes NA    NA NA NA NA
## 381           Yes NA    NA NA NA NA
## 382           Yes NA    NA NA NA NA
## 383           Yes NA    NA NA NA NA
## 384           Yes NA    NA NA NA NA
## 385           Yes NA    NA NA NA NA
## 386           Yes NA    NA NA NA NA
## 387            No NA    NA NA NA NA
## 388            No NA    NA NA NA NA
## 389           Yes NA    NA NA NA NA
## 390            No NA    NA NA NA NA
## 391           Yes NA    NA NA NA NA
## 392            No NA    NA NA NA NA
## 393           Yes NA    NA NA NA NA
## 394            No NA    NA NA NA NA
## 395           Yes NA    NA NA NA NA
## 396           Yes NA    NA NA NA NA
## 397           Yes NA    NA NA NA NA
## 398            No NA    NA NA NA NA
## 399            No NA    NA NA NA NA
## 400            No NA    NA NA NA NA
## 401           Yes NA    NA NA NA NA
## 402           Yes NA    NA NA NA NA
## 403            No NA    NA NA NA NA
## 404            No NA    NA NA NA NA
## 405            No NA    NA NA NA NA
## 406            No NA    NA NA NA NA
## 407           Yes NA    NA NA NA NA
## 408            No NA    NA NA NA NA
## 409            No NA    NA NA NA NA
## 410           Yes NA    NA NA NA NA
```

```
summary(Professional)
```

```
##       age           gender           real_estate         investments
##  Min.   :19.00   Length:410         Length:410         Min.   :     0
##  1st Qu.:28.00   Class :character   Class :character   1st Qu.: 18300
##  Median :30.00   Mode  :character   Mode  :character   Median : 24800
##  Mean   :30.11                                         Mean   : 28538
##  3rd Qu.:33.00                                         3rd Qu.: 34275
##  Max.   :42.00                                         Max.   :133400
##    num_trans      has_broadband          income       have_children
##  Min.   : 0.000   Length:410         Min.   : 16200   Length:410
##  1st Qu.: 4.000   Class :character   1st Qu.: 51625   Class :character
##  Median : 6.000   Mode  :character   Median : 66050   Mode  :character
##  Mean   : 5.973                      Mean   : 74460
##  3rd Qu.: 7.000                      3rd Qu.: 88775
##  Max.   :21.000                      Max.   :322500
##     NA                NA                 NA              NA              NA
##  Mode:logical   Length:410         Mode:logical   Mode:logical   Mode:logical
##  NA's:410       Class :character   NA's:410       NA's:410       NA's:410
##                 Mode  :character
##
##
##
##     NA
##  Mode:logical
##  NA's:410
##
##
##
##
```

```
#b. Develop 95% confidence intervals for the mean age and household income of subscribers.
# 95%  age
age_mean <- mean(Professional$age)
age_sd <- sd(Professional$age)
n <- length(Professional$age)
age_ci <- t.test(Professional$age)$conf.int
age_ci
```

```
## [1] 29.72153 30.50286
## attr(,"conf.level")
## [1] 0.95
```

```
# 95% income
income_mean <- mean(Professional$income)
income_sd <- sd(Professional$income)
income_ci <- t.test(Professional$income)$conf.int
income_ci
```

```
## [1] 71079.26 77839.77
## attr(,"conf.level")
## [1] 0.95
```

```
#c. Develop 95% confidence intervals for the proportion of subscribers who have broadband access at home and the proportion of subscribers who have children.
# 95% has_broadband
broadband_test <- prop.test(sum(Professional$has_broadband=="Yes"), length(Professional$has_broadband), conf.level = 0.95)
print(broadband_test)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  sum(Professional$has_broadband == "Yes") out of length(Professional$has_broadband), null probability 0.5
## X-squared = 24.88, df = 1, p-value = 6.1e-07
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.5753252 0.6710862
## sample estimates:
##         p
## 0.6243902
```

```
# 95% have_children
children_test <- prop.test(sum(Professional$have_children=="Yes"), length(Professional$have_children), conf.level = 0.95)
print(children_test)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  sum(Professional$have_children == "Yes") out of length(Professional$have_children), null probability 0.5
## X-squared = 1.778, df = 1, p-value = 0.1824
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.4845521 0.5830908
## sample estimates:
##         p
## 0.5341463
```

```
#d. Would Young Professional be a good advertising outlet for online brokers? Justify your conclusion with statistical data.
#yes real_estate>0.5
real_estate <- sum(Professional$real_estate=="Yes")/length(Professional$real_estate)
real_estate
```

```
## [1] 0.4414634
```

```
#e. Would this magazine be a good place to advertise for companies selling educational sof
tware and computer games for young children?
# yes,have_children>0.5
have_children <- sum(Professional$have_children=="Yes")/length(Professional$have_children)
have_children
```

```
## [1] 0.5341463
```

```
#f. Comment on the types of articles you believe would be of interest to readers of Young
Professional.

# eg:children estate
```

Question #5: Quality Associate, Inc. (Attached Data: Quality)

```
Quality <- read.csv("data/Quality.csv", header = TRUE, sep = ",")
# a. Conduct a hypothesis test for each sample at the .01 level of significance and determ
ine what action, if any, should be taken. Provide the p-value for each test.
alpha <- 0.01
t_test_results <- lapply(Quality[, 1:4], function(x) t.test(x, mu = 12, var.equal = TRUE))
p_values <- sapply(t_test_results, function(test) test$p.value)
names(p_values) <- c("Sample 1", "Sample 2", "Sample 3", "Sample 4")
p_values
```

```
##    Sample 1    Sample 2    Sample 3    Sample 4
## 0.312729582 0.481820940 0.006468822 0.039058947
```

```
# b. compute the standard deviation for each of the four samples. does the assumption of .
21 for the population standard deviation appear reasonable?
# is reasonable
sample_sds <- sapply(Quality[, 1:4], sd)
names(sample_sds) <- c("Sample 1", "Sample 2", "Sample 3", "Sample 4")
sample_sds
```

```
##  Sample 1  Sample 2  Sample 3  Sample 4
## 0.2203560 0.2203560 0.2071706 0.2061090
```

```
# c. compute limits for the sample mean x around miu= 12 such that, as long as a new sampl
e mean is within those limits, the process will be considered to be operating satisfactori
ly. if x exceeds the upper limit or if x is below the lower limit, corrective action will
be taken. these limits are referred to as upper and lower control limits for quality contr
ol purposes.
miu = 12
sigma = 0.21
n = 30
alpht = 0.01
zinterval <- function(miu,sigma,prob,n) {
  return(c(miu + qnorm(prob) * sigma / sqrt(n),
           miu - qnorm(prob) * sigma / sqrt(n)))}
zinterval(12,0.21,0.01,30)
```

```
## [1] 11.91081 12.08919
```

```
# d. discuss the implications of changing the level of significance to a larger value. wha
t mistake or error could increase if the level of significance is increased?
#Type I error will increase
zinterval(12,0.21,0.05,30)
```

```
## [1] 11.93694 12.06306
```

```
#Type I error: The probability of incorrectly rejecting the true null hypothesis increases
. This can lead to unnecessary corrective actions.
#Type II error: The probability of incorrectly failing to reject a false null hypothesis i
s reduced because it is easier to reject the null hypothesis.
```

Question #6: Vacation occupancy rates were expected to be up during March 2008 in Myrtle

Beach, South Carolina (the sun news, February 29, 2008). Data in the file Occupancy (Attached file Occupancy) will allow you to replicate the findings presented in the newspaper. The data show units rented and not rented for a random sample of vacation properties during the first week of March 2007 and March 2008.

```
Occupancy <- read.csv("data/Occupancy.csv", skip = 1)
Occupancy
```

```
##    March.2007 March.2008
## 1         Yes         No
## 2          No        Yes
## 3         Yes        Yes
## 4          No         No
## 5          No        Yes
## 6         Yes         No
## 7          No         No
## 8          No        Yes
## 9          No        Yes
## 10        Yes        Yes
## 11         No         No
## 12        Yes         No
## 13         No        Yes
## 14         No        Yes
## 15         No        Yes
## 16         No        Yes
## 17         No         No
## 18         No        Yes
## 19         No        Yes
## 20         No        Yes
## 21         No        Yes
## 22         No         No
## 23        Yes        Yes
## 24        Yes         No
## 25        Yes        Yes
## 26        Yes         No
## 27         No        Yes
## 28        Yes         No
## 29         No         No
## 30         No         No
## 31         No        Yes
## 32        Yes        Yes
## 33         No         No
## 34        Yes        Yes
## 35         No         No
## 36        Yes         No
## 37         No        Yes
## 38        Yes        Yes
## 39        Yes         No
## 40         No         No
## 41         No        Yes
## 42        Yes        Yes
## 43         No         No
## 44         No        Yes
## 45        Yes         No
## 46        Yes         No
## 47         No        Yes
## 48         No        Yes
## 49         No        Yes
## 50         No        Yes
## 51         No         No
## 52         No         No
## 53        Yes         No
```

```
## 54          No         No
## 55          No        Yes
## 56          No         No
## 57          No         No
## 58          No         No
## 59         Yes         No
## 60          No        Yes
## 61         Yes         No
## 62          No        Yes
## 63          No        Yes
## 64         Yes         No
## 65          No         No
## 66          No         No
## 67          No         No
## 68         Yes         No
## 69          No        Yes
## 70          No         No
## 71         Yes         No
## 72         Yes         No
## 73         Yes         No
## 74          No         No
## 75          No        Yes
## 76          No        Yes
## 77         Yes        Yes
## 78          No         No
## 79          No         No
## 80          No        Yes
## 81          No        Yes
## 82          No        Yes
## 83         Yes         No
## 84          No         No
## 85         Yes        Yes
## 86          No         No
## 87         Yes        Yes
## 88          No         No
## 89         Yes         No
## 90          No        Yes
## 91          No        Yes
## 92          No        Yes
## 93         Yes        Yes
## 94          No        Yes
## 95          No        Yes
## 96         Yes        Yes
## 97          No        Yes
## 98          No        Yes
## 99          No         No
## 100         No         No
## 101        Yes        Yes
## 102         No         No
## 103        Yes        Yes
## 104        Yes         No
## 105        Yes         No
## 106         No         No
## 107         No         No
## 108        Yes        Yes
```

```
## 109          No        Yes
## 110         Yes         No
## 111         Yes        Yes
## 112         Yes         No
## 113          No         No
## 114         Yes        Yes
## 115          No         No
## 116         Yes         No
## 117          No         No
## 118         Yes         No
## 119         Yes         No
## 120          No         No
## 121          No         No
## 122          No         No
## 123          No         No
## 124          No        Yes
## 125          No        Yes
## 126          No         No
## 127          No        Yes
## 128          No         No
## 129         Yes         No
## 130         Yes         No
## 131         Yes        Yes
## 132          No         No
## 133          No        Yes
## 134          No        Yes
## 135         Yes         No
## 136          No        Yes
## 137          No         No
## 138         Yes         No
## 139         Yes         No
## 140          No         No
## 141         Yes        Yes
## 142         Yes        Yes
## 143          No        Yes
## 144         Yes         No
## 145          No        Yes
## 146         Yes        Yes
## 147          No        Yes
## 148         Yes         No
## 149         Yes         No
## 150          No         No
## 151          No
## 152          No
## 153         Yes
## 154          No
## 155          No
## 156          No
## 157          No
## 158          No
## 159          No
## 160          No
## 161          No
## 162          No
## 163         Yes
```

```
## 164        Yes
## 165         No
## 166         No
## 167         No
## 168        Yes
## 169         No
## 170         No
## 171         No
## 172         No
## 173         No
## 174         No
## 175         No
## 176        Yes
## 177        Yes
## 178         No
## 179         No
## 180        Yes
## 181         No
## 182         No
## 183        Yes
## 184         No
## 185        Yes
## 186         No
## 187         No
## 188        Yes
## 189        Yes
## 190         No
## 191         No
## 192         No
## 193         No
## 194         No
## 195        Yes
## 196        Yes
## 197         No
## 198         No
## 199         No
## 200         No
```

```
#a. Estimate the proportion of units rented during the first week of March 2007 and the fi
rst week of March 2008.
#sum(Occupancy$March.2007 %in% c("Yes"))/200
#sum(Occupancy$March.2008 %in% c("Yes"))/150
rented2007 <- sum(Occupancy$March.2007=="Yes")/length(Occupancy$March.2007)
rented2007
```

```
## [1] 0.35
```

```
rented2008 <- sum(Occupancy$March.2008=="Yes")/150
rented2008
```

```
## [1] 0.4666667
```

```
# b. Provide a 95% confidence interval for the difference in proportions.
pa <- sum(Occupancy$March.2007 %in% c("Yes"))/200
pb <- sum(Occupancy$March.2008 %in% c("Yes"))/150
e <- qnorm(0.975) * sqrt(pa*(1-pa)/200 + pb*(1-pb)/150)
c(pa-pb-e,pa-pb+e)
```

```
## [1] -0.22031818 -0.01301516
```

```
# c. On the basis of your findings, does it appear March rental rates for 2008 will be up
from those a year earlier?
# -0.22031818 -0.01301516 doesn't include Zero,Impossible to judge
```

Question #7: Air Force Training Program (data file: Training)

```
Training <- read.csv("data/Training.csv", header = TRUE, sep = ",")
Training
```

```
##    Current Proposed
## 1       76       74
## 2       76       75
## 3       77       77
## 4       74       78
## 5       76       74
## 6       74       80
## 7       74       73
## 8       77       73
## 9       72       78
## 10      78       76
## 11      73       76
## 12      78       74
## 13      75       77
## 14      80       69
## 15      79       76
## 16      72       75
## 17      69       72
## 18      79       75
## 19      72       72
## 20      70       76
## 21      70       72
## 22      81       77
## 23      76       73
## 24      78       77
## 25      72       69
## 26      82       77
## 27      72       75
## 28      73       76
## 29      71       74
## 30      70       77
## 31      77       75
## 32      78       78
## 33      73       72
## 34      79       77
## 35      82       78
## 36      65       78
## 37      77       76
## 38      79       75
## 39      73       76
## 40      76       76
## 41      81       75
## 42      69       76
## 43      75       80
## 44      75       77
## 45      77       76
## 46      79       75
## 47      76       73
## 48      78       77
## 49      76       77
## 50      76       77
## 51      73       79
## 52      77       75
## 53      84       75
```

```
## 54      74       72
## 55      74       82
## 56      69       76
## 57      79       76
## 58      66       74
## 59      70       72
## 60      74       78
## 61      72       71
```

```
#a. use appropriate descriptive statistics to summarize the training time data for each me
thod.what similarities or differences do you observe from the sample data?
skimr::skim(Training)
```

Data summary

Name	Training
Number of rows	61
Number of columns	2
_______________________	
Column type frequency:	
numeric	2
________________________	
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Current	0	1	75.07	3.94	65	72	76	78	84	▁▅▇▆▂
Proposed	0	1	75.43	2.51	69	74	76	77	82	▁▆▇▆▁

```
summary(Training$Current)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   65.00   72.00   76.00   75.07   78.00   84.00
```

```
summary(Training$Proposed)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   69.00   74.00   76.00   75.43   77.00   82.00
```

```
#b. Comment on any difference between the population means for the two methods. Discuss yo
ur findings.
Current <- mean(Training$Current)
Current
```

```
## [1] 75.06557
```

```
Proposed <- mean(Training$Proposed)
Proposed
```

```
## [1] 75.42623
```

```
#t.test(Training$Current,Training$Proposed)
# Proposed is Bigger

# c. compute the standard deviation and variance for each training method. conduct a hypot
hesis test about the equality of population variances for the two training methods. Discus
s your findings.

sdCurrent <- sd(Training$Current)
varCurrent <- var(Training$Current)
sdProposed <- sd(Training$Proposed)
varProposed <- var(Training$Proposed)
sdCurrent
```

```
## [1] 3.944907
```

```
varCurrent
```

```
## [1] 15.5623
```

```
sdProposed
```

```
## [1] 2.506385
```

```
varProposed
```

```
## [1] 6.281967
```

```
# 方差相等性检验（F-test）
vartest <- var.test(Training$Current, Training$Proposed)
vartest
```

```
##
##  F test to compare two variances
##
## data:  Training$Current and Training$Proposed
## F = 2.4773, num df = 60, denom df = 60, p-value = 0.000578
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.486267 4.129135
## sample estimates:
## ratio of variances
##           2.477296
```

```
# d. what conclusion can you reach about any differences between the two methods? what is your recommendation? explain.
if (Proposed < Current && vartest$p.value > 0.05) {
  cat("The current method is preferred\n")
} else {
  cat("The proposed method is preferred\n")
}
```

```
## The proposed method is preferred
```

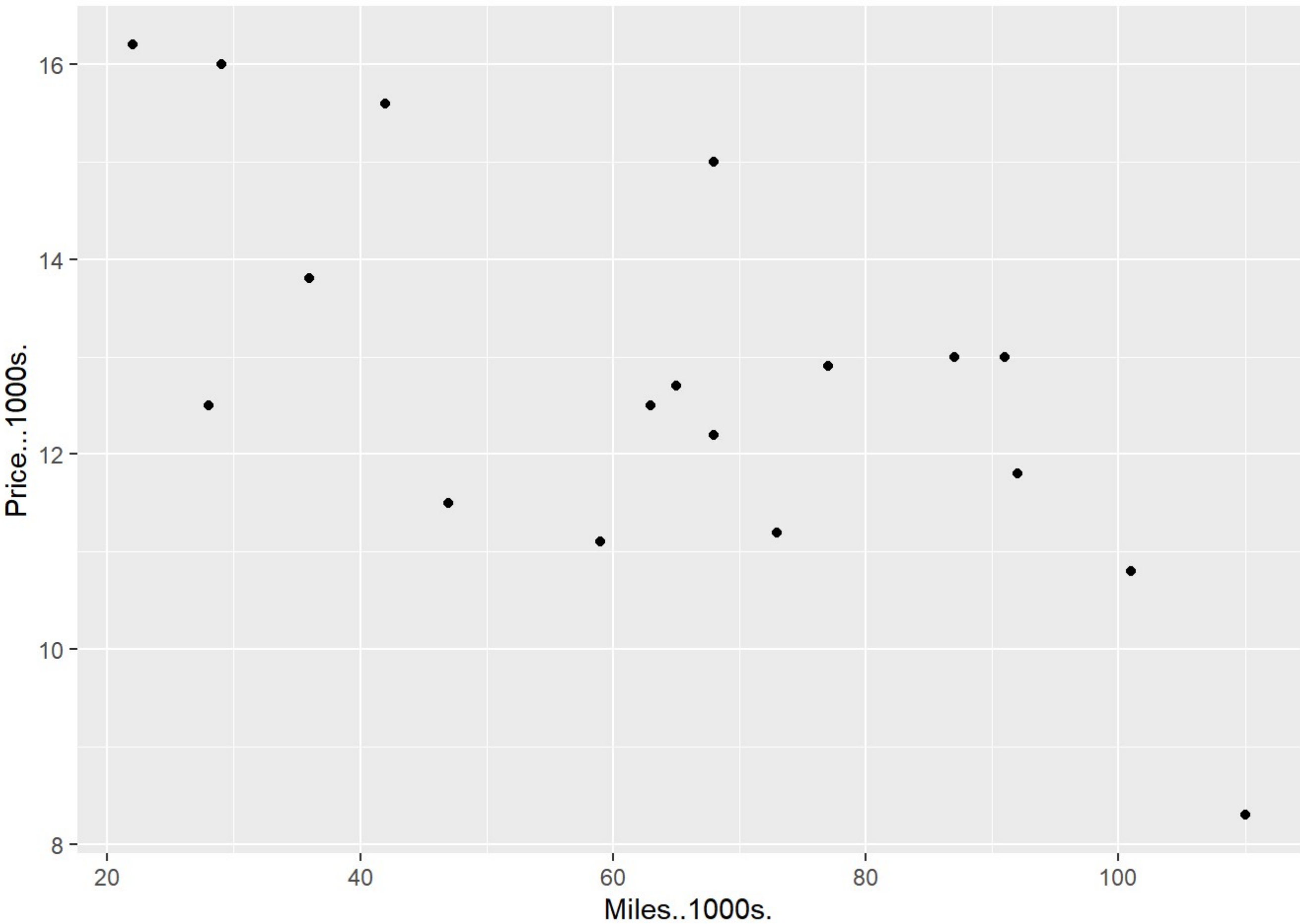
```
#e. can you suggest other data or testing that might be desirable before making a final decision on the training program to be used in the future
# such as student satisfaction
```

Question #8: The Toyota Camry is one of the best-selling cars in North America. The cost of a previously owned Camry depends upon many factors, including the model year, mileage, and condition. To investigate the relationship between the car's mileage and the sales price for a 2007 model year Camry, Attached data file Camry show the mileage and sale price for 19 sales (Pricehub website, February 24, 2012).

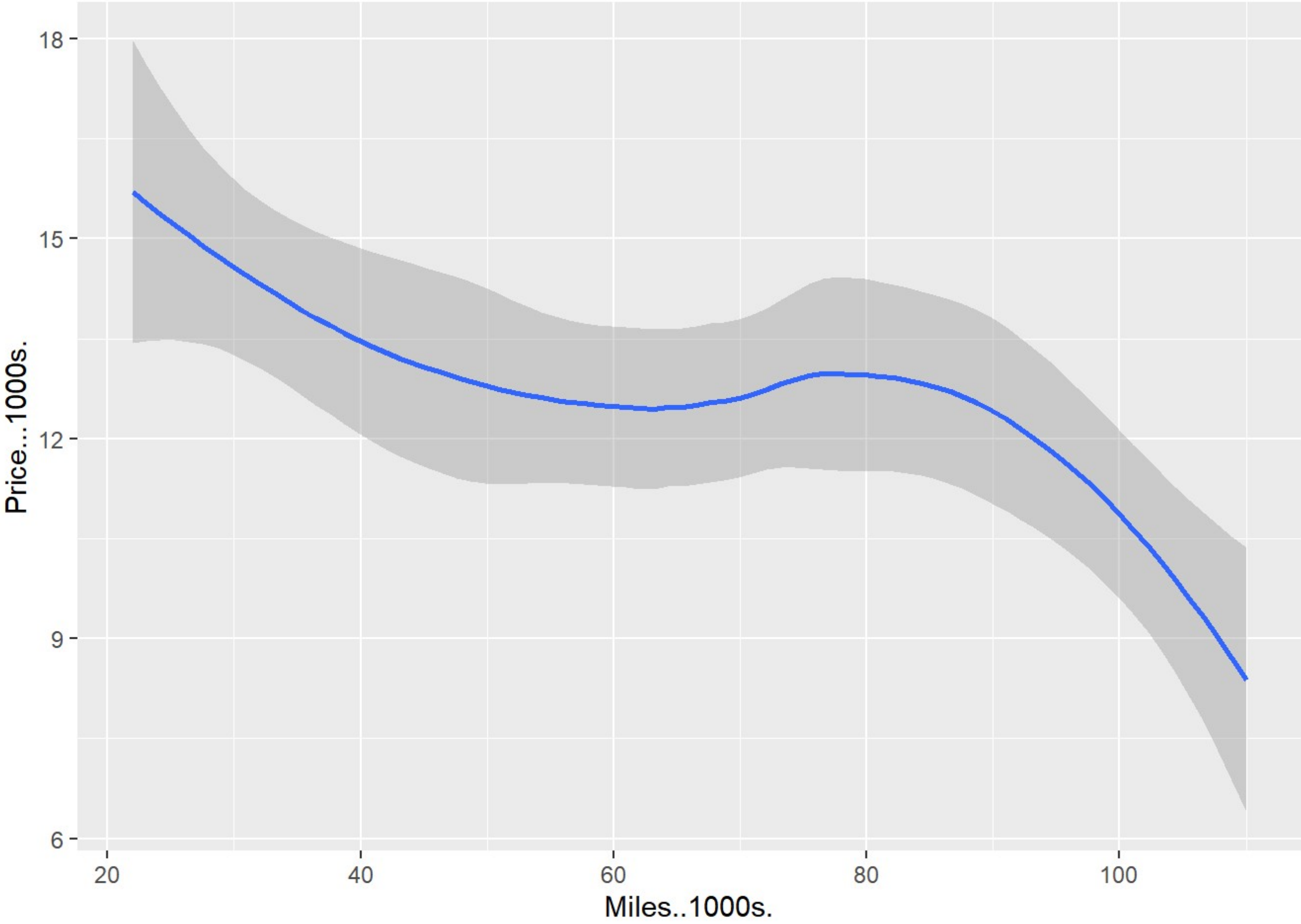
```
Camry <- read.csv("data/Camry.csv", header = TRUE, sep = ",")
Camry
```

```
##    Miles..1000s. Price...1000s.
## 1             22           16.2
## 2             29           16.0
## 3             36           13.8
## 4             47           11.5
## 5             63           12.5
## 6             77           12.9
## 7             73           11.2
## 8             87           13.0
## 9             92           11.8
## 10           101           10.8
## 11           110            8.3
## 12            28           12.5
## 13            59           11.1
## 14            68           15.0
## 15            68           12.2
## 16            91           13.0
## 17            42           15.6
## 18            65           12.7
## 19           110            8.3
```

```
#a. Develop a scatter diagram with the car mileage on the horizontal axis and the price on
the vertical axis.
ggplot(Camry, aes(Miles..1000s., Price...1000s.)) +
  geom_point()
```



```
# b. what does the scatter diagram developed in part (a) indicate about the relationship b
etween the two variables?
ggplot(Camry, aes(Miles..1000s., Price...1000s.)) +
  geom_smooth(method = 'loess',formula = 'y ~ x')
```




```
# decrease
# c. Develop the estimated regression equation that could be used to predict the price ($1
000s) given the miles (1000s).
lm_camry <- lm(Price...1000s.~ Miles..1000s., data = Camry)
summary(lm_camry)
```

```
##
## Call:
## lm(formula = Price...1000s. ~ Miles..1000s., data = Camry)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.32408 -1.34194  0.05055  1.12898  2.52687
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   16.46976    0.94876  17.359 2.99e-12 ***
## Miles..1000s. -0.05877    0.01319  -4.455 0.000348 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.541 on 17 degrees of freedom
## Multiple R-squared:  0.5387, Adjusted R-squared:  0.5115
## F-statistic: 19.85 on 1 and 17 DF,  p-value: 0.0003475
```

```
# d. Test for a significant relationship at the .05 level of significance.

#p-value = 0.0003475<0.05

# e. Did the estimated regression equation provide a good fit? Explain.

# R-squared:  0.5387 good fig

# f. Provide an interpretation for the slope of the estimated regression equation.

# estimated -0.05877  Miles increase Price decrease

# g. Suppose that you are considering purchasing a previously owned 2007 Camry that has be
en driven 60,000 miles. Using the estimated regression equation developed in part (c), pre
dict the price for this car. Is this the price you would offer the seller.

# Price = 16.470 - 0.059 * miles  12942
```

Question #9

```
WE <- read_excel("data/WE.xlsx") %>%
  set_names("id","churn","happy_index","chg_hi","support","chg_supprt",
            "priority","chg_priority","log_in_fre","chg_blog_fre","chg_vis","y_age","chg_i
nterval")
list_name = c("id","churn","happy_index","chg_hi","support","chg_supprt",
            "priority","chg_priority","log_in_fre","chg_blog_fre","chg_vis","y_age","chg_i
nterval")
WE
```

```
## # A tibble: 6,347 × 13
##       id churn happy_index chg_hi support chg_supprt priority chg_priority
##    <dbl> <dbl>       <dbl>  <dbl>   <dbl>      <dbl>    <dbl>        <dbl>
##  1     1     0           0      0       0          0        0            0
##  2     2     0          62      4       0          0        0            0
##  3     3     0           0      0       0          0        0            0
##  4     4     0         231      1       1         -1        3            0
##  5     5     0          43     -1       0          0        0            0
##  6     6     0         138    -10       0          0        0            0
##  7     7     0         180     -5       1          1        3            3
##  8     8     0         116    -11       0          0        0            0
##  9     9     0          78     -7       1         -2        3            0
## 10    10     0          78    -37       0          0        0            0
## # ℹ 6,337 more rows
## # ℹ 5 more variables: log_in_fre <dbl>, chg_blog_fre <dbl>, chg_vis <dbl>,
## #   y_age <dbl>, chg_interval <dbl>
```

```
#a
WE %>%
  select(-id) %>%
  group_by(churn) %>%
  group_modify(~{
    .x %>%
      purrr::map_dfc(mean, na.rm = TRUE)
  }) %>%
  ungroup()
```

```
## # A tibble: 2 × 12
##   churn happy_index chg_hi support chg_supprt priority chg_priority log_in_fre
##   <dbl>       <dbl>  <dbl>   <dbl>      <dbl>    <dbl>        <dbl>      <dbl>
## 1     0        88.6   5.53   0.724   -0.00930    0.830       0.0327      16.1
## 2     1        63.3  -3.74   0.372    0.0372     0.500      -0.0167       8.06
## # ℹ 4 more variables: chg_blog_fre <dbl>, chg_vis <dbl>, y_age <dbl>,
## #   chg_interval <dbl>
```

```
#b.
compare_means <- function(var_name) {
  group_0 <- WE[WE$churn == 0, var_name]
  group_1 <- WE[WE$churn == 1, var_name]
  result <- t.test(group_0, group_1)
  return(result)
}

for(i in 3:length(list_name)){
  print(list_name[i])
  print(compare_means(list_name[i]))
}
```

```
## [1] "happy_index"
## 
##  Welch Two Sample t-test
## 
## data:  group_0 and group_1
## t = 7.6242, df = 369.36, p-value = 2.097e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  18.79956 31.86737
## sample estimates:
## mean of x mean of y 
##  88.60591  63.27245 
## 
## [1] "chg_hi"
## 
##  Welch Two Sample t-test
## 
## data:  group_0 and group_1
## t = 5.7835, df = 365.71, p-value = 1.571e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.116137 12.417972
## sample estimates:
## mean of x mean of y 
##  5.530212 -3.736842 
## 
## [1] "support"
## 
##  Welch Two Sample t-test
## 
## data:  group_0 and group_1
## t = 5.5099, df = 419.22, p-value = 6.281e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2269082 0.4785969
## sample estimates:
## mean of x mean of y 
## 0.7242696 0.3715170 
## 
## [1] "chg_supprt"
## 
##  Welch Two Sample t-test
## 
## data:  group_0 and group_1
## t = -0.63198, df = 406.9, p-value = 0.5278
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.19092606  0.09803036
## sample estimates:
##    mean of x    mean of y 
## -0.009296149  0.037151703 
## 
## [1] "priority"
## 
```

```
##  Welch Two Sample t-test
##
## data:  group_0 and group_1
## t = 5.1428, df = 373.13, p-value = 4.381e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2038355 0.4562009
## sample estimates:
## mean of x mean of y
## 0.8295759 0.4995577
##
## [1] "chg_priority"
##
##  Welch Two Sample t-test
##
## data:  group_0 and group_1
## t = 0.64116, df = 364.49, p-value = 0.5218
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1020692  0.2008252
## sample estimates:
##   mean of x   mean of y
##  0.03268184 -0.01669615
##
## [1] "log_in_fre"
##
##  Welch Two Sample t-test
##
## data:  group_0 and group_1
## t = 3.5709, df = 362.67, p-value = 0.0004037
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.628884 12.525166
## sample estimates:
## mean of x mean of y
##  16.13894   8.06192
##
## [1] "chg_blog_fre"
##
##  Welch Two Sample t-test
##
## data:  group_0 and group_1
## t = 2.5315, df = 695.95, p-value = 0.01158
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.06133902 0.48529282
## sample estimates:
##  mean of x  mean of y
##  0.1711487 -0.1021672
##
## [1] "chg_vis"
##
##  Welch Two Sample t-test
##
## data:  group_0 and group_1
```

```
## t = 1.9136, df = 448, p-value = 0.05631
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -5.463729 410.218457
## sample estimates:
## mean of x mean of y
##  106.6096  -95.7678
##
## [1] "y_age"
##
##  Welch Two Sample t-test
##
## data:  group_0 and group_1
## t = -2.9811, df = 379.9, p-value = 0.003057
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.5461200 -0.5223121
## sample estimates:
## mean of x mean of y
##  18.81873  20.35294
##
## [1] "chg_interval"
##
##  Welch Two Sample t-test
##
## data:  group_0 and group_1
## t = -4.0971, df = 346.03, p-value = 5.215e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.362712 -2.586515
## sample estimates:
## mean of x mean of y
##  3.511454  8.486068
```

```
#c.

model_q9 <- glm(churn ~., data = WE[, c("churn", list_name)], family = binomial())
summary(model_q9)
```

```
##
## Call:
## glm(formula = churn ~ ., family = binomial(), data = WE[, c("churn",
##     list_name)])
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.529e-01  2.275e-01  -3.309 0.000935 ***
## id           -4.102e-04  3.991e-05 -10.279  < 2e-16 ***
## happy_index  -7.710e-03  1.250e-03  -6.166    7e-10 ***
## chg_hi       -7.057e-03  2.532e-03  -2.787 0.005317 **
## support      -1.098e-01  1.032e-01  -1.064 0.287402
## chg_supprt    1.297e-01  8.958e-02   1.447 0.147803
## priority      3.486e-04  1.018e-01   0.003 0.997267
## chg_priority -3.007e-02  7.801e-02  -0.385 0.699910
## log_in_fre    5.686e-04  2.209e-03   0.257 0.796865
## chg_blog_fre  3.361e-03  2.086e-02   0.161 0.872005
## chg_vis      -9.431e-05  4.223e-05  -2.233 0.025547 *
## y_age        -2.324e-02  6.495e-03  -3.579 0.000346 ***
## chg_interval  1.075e-02  3.686e-03   2.916 0.003542 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2553.1  on 6346  degrees of freedom
## Residual deviance: 2341.6  on 6334  degrees of freedom
## AIC: 2367.6
##
## Number of Fisher Scoring iterations: 7
```

```
# d
not_churned_data <- WE[WE$churn == 1, ]

probabilities_q9 <- predict(model_q9, newdata = not_churned_data, type = "response")
not_churned_data$Loss_probability <- probabilities_q9
sorted_data <- not_churned_data[order(not_churned_data$Loss_probability), ]

top_100_user_ids <- sorted_data$id[1:100]
print(top_100_user_ids)
```

```
##   [1] 5294 2704 5271 4455 4740 3984 4817 5560 4899 4793 4178 5587   54 4642 1892
##  [16] 5350 5422 4239 4137 5765 4055 4299 4489 1764 3849 2869 4925 4174 3005 2980
##  [31] 3134 4900 4313 2578 2212 1037 2506 3558 1641 3670 4006 3494 5405 3806 4058
##  [46] 5214 3135 6263 1810 2682 3626 6127 2073 3876 3064 3886 2249 3492 3014 3069
##  [61] 1489 4363 2941 4483 1339 1997 4885 2701 3779  978 2650 2616 1043 4459 3682
##  [76] 1895 2357 4185 4629 1067  886 1905 2011 4157 1551 1992 2693 3256 1947 1872
##  [91] 5426 2628  875 4582 2489 4020 1724 4533 5052  153
```