Demo Independent t-test

Example 1: You are a HR Manager for a large corporation and have been interested in implementing a new training program for your marketing employees. To test this, you have 80 new employees that are about to be trained on marketing strategies. 40 of the employees you will put through your typical 1-day training program (Group A), the other 40 employees will be placed in your new week-long training program (Group B). You ask your assistant to randomly take 40 names out of a hat to determine which group gets put into the new program. You decide to test which training program works better by the amount of marketing ideas produced after the first month of being in the company. Based on your findings, should you implement the new training program or not? Why?

Answer

> Based on the results of the statistical test, there is a significantly higher difference ($p < .001$) between Group B ($M = 38.1$) and Group A ($M = 29.6$). All assumptions underlying independent samples t-test have been met (independence of observations, approximately normally distributed, homogeneity of variance [Levene's test $p = .134$]). As such, it is recommended to implement the new training program, as the new group created a significantly higher number of Marketing Ideas.

```
library(psych)
library(jmv)
```

```
##
## Attaching package: 'jmv'

## The following object is masked from 'package:psych':
##
##     pca

## The following object is masked from 'package:stats':
##
##     anova
```

```
library(summarytools)
```

IMPORT DATA

- View(dat_A)

```
dat_A <- read.csv("https://www.dropbox.com/s/mhcbxwwb6rofcuy/independent.A.csv?dl=1")
```

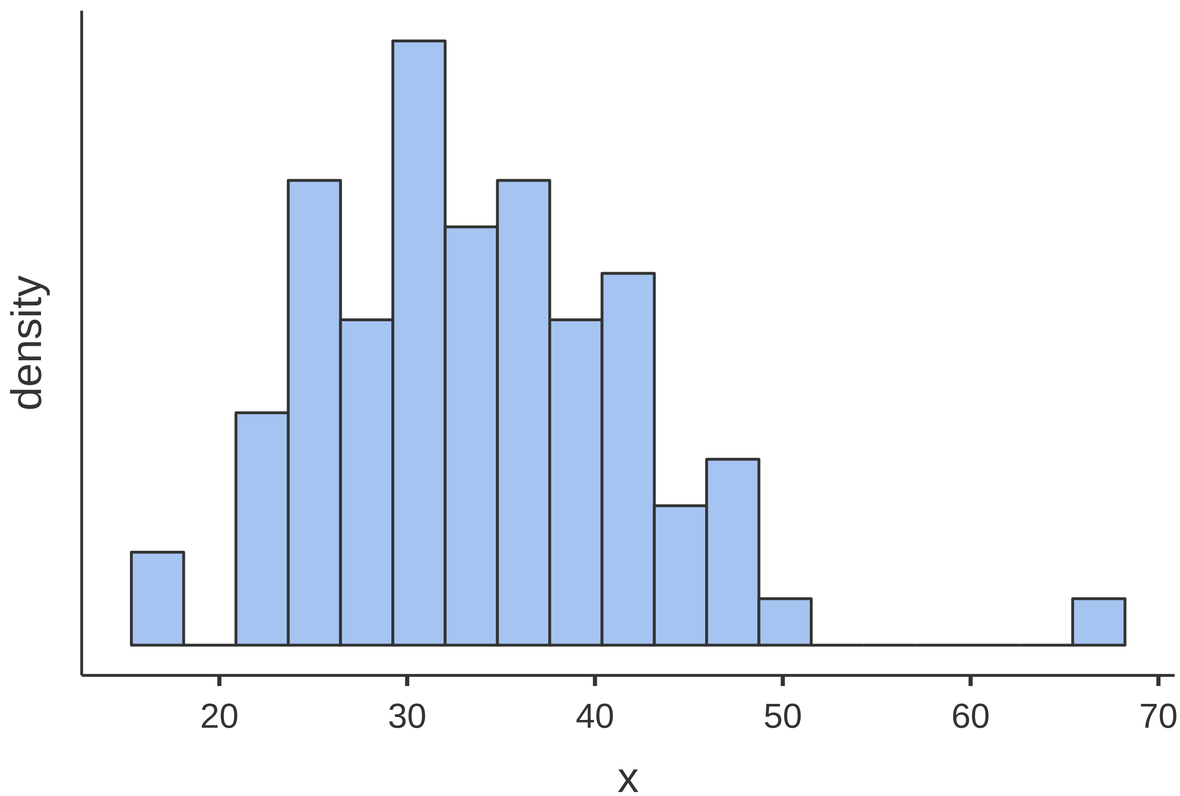
DESCRIPTIVES

Not split by group

We need it to be by group. Here the descriptives lump all the data in the data frame together * class(desc) —> What type of object is this? * class(dat_A) —> data.frame

```
desc <- descriptives(dat_A, vars = c('Marketing_Ideas'), hist = TRUE, sd = TRUE, se = TRUE, skew = TRUE
desc
```

```
##
## DESCRIPTIVES
##
## Descriptives
## ------------------------------------------
##                          Marketing_Ideas
## ------------------------------------------
##   N                                   80
##   Missing                              0
##   Mean                              33.8
##   Std. error mean                  0.922
##   Median                            33.2
##   Standard deviation                8.24
##   Minimum                           17.2
##   Maximum                           67.3
##   Skewness                         0.796
##   Std. error skewness              0.269
##   Kurtosis                          2.22
##   Std. error kurtosis              0.532
## ------------------------------------------
```

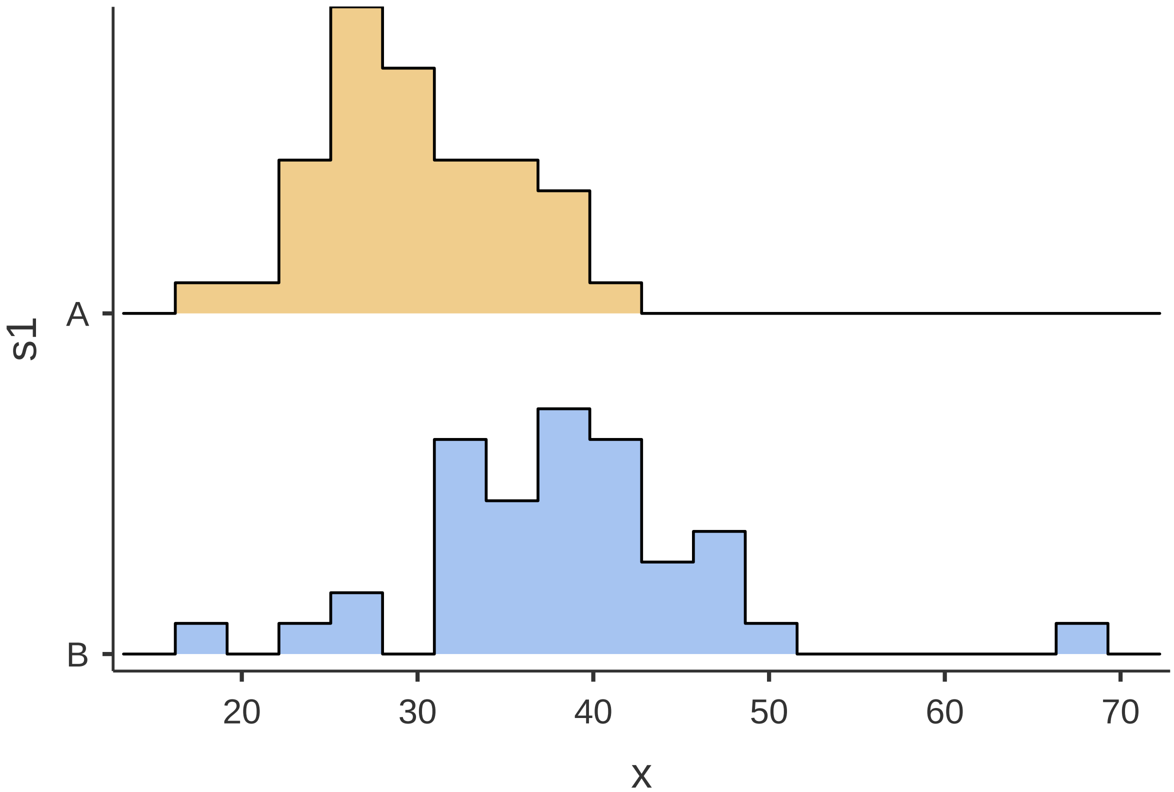


Split by group

We need this. * View(groupdesc)

```r
groupdesc <- descriptives(dat_A, vars = c('Marketing_Ideas'), splitBy = 'Group', hist = TRUE, sd = TRUE
groupdesc
```

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  ---------------------------------------------------
##                           Group    Marketing_Ideas
##  ---------------------------------------------------
##    N                      A                     40
##                           B                     40
##    Missing                A                      0
##                           B                      0
##    Mean                   A                   29.6
##                           B                   38.1
##    Std. error mean        A                  0.855
##                           B                   1.33
##    Median                 A                   29.3
##                           B                   37.8
##    Standard deviation     A                   5.41
##                           B                   8.42
##    Minimum                A                   17.2
##                           B                   17.9
##    Maximum                A                   40.6
##                           B                   67.3
##    Skewness               A                  0.135
##                           B                  0.576
##    Std. error skewness    A                  0.374
##                           B                  0.374
##    Kurtosis               A                 -0.372
##                           B                   3.07
##    Std. error kurtosis    A                  0.733
##                           B                  0.733
##  ---------------------------------------------------
```




```
dat_A$Group <- as.factor(dat_A$Group)
```

ANALYSIS

independent samples t-test

{jmv library} * vars = DV, group = IV, eqv = homogeneity of variance

```
ttestIS(data = dat_A, vars = 'Marketing_Ideas', group = 'Group', eqv = TRUE, effectSize = TRUE, ci = TRU
```

```
##
##  INDEPENDENT SAMPLES T-TEST
##
##  Independent Samples T-Test
##  ----------------------------------------------------------------------------------------------------
##                                       statistic    df      p          Lower     Upper     Cohen's d
##  ----------------------------------------------------------------------------------------------------
##    Marketing_Ideas    Student's t         -5.40    78.0    < .001     -11.7     -5.39         -1.21
##  ----------------------------------------------------------------------------------------------------
##
##
##  ASSUMPTIONS
##
##  Test of Equality of Variances (Levene's)
##  ------------------------------------------
```

```
##                        F       df    p
##  ------------------------------------------
##    Marketing_Ideas    2.29     1    0.134
##  ------------------------------------------
##    Note. A low p-value suggests a
##    violation of the assumption of equal
##    variances
##
##
##  Group Descriptives
##  ---------------------------------------------------------------------
##                       Group    N     Mean    Median    SD      SE
##  ---------------------------------------------------------------------
##    Marketing_Ideas    A        40    29.6      29.3    5.41    0.855
##                       B        40    38.1      37.8    8.42     1.33
##  ---------------------------------------------------------------------
```

VISUALIZATION

```r
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```r
library(plyr)
```

GGPLOT FUNCTION

```r
# creating this function will allow you to generate summary stats needed for visualization - mean score

summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE,
                      conf.interval=.95, .drop=TRUE) {
  library(plyr)

  # New version of length which can handle NA's: if na.rm==T, don't count them
  length2 <- function (x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else       length(x)
  }

  # This does the summary. For each group's data frame, return a vector with
  # N, mean, and sd
  datac <- ddply(data, groupvars, .drop=.drop,
                 .fun = function(xx, col) {
                   c(N    = length2(xx[[col]], na.rm=na.rm),
                     mean = mean   (xx[[col]], na.rm=na.rm),
                     sd   = sd     (xx[[col]], na.rm=na.rm)
```

```
                   )
                 },
                 measurevar
  )

  # Rename the "mean" column
  datac <- rename(datac, c("mean" = measurevar))

  datac$se <- datac$sd / sqrt(datac$N)  # Calculate standard error of the mean

  # Confidence interval multiplier for standard error
  # Calculate t-statistic for confidence interval:
  # e.g., if conf.interval is .95, use .975 (above/below), and use df=N-1
  ciMult <- qt(conf.interval/2 + .5, datac$N-1)
  datac$ci <- datac$se * ciMult

  return(datac)
}
```

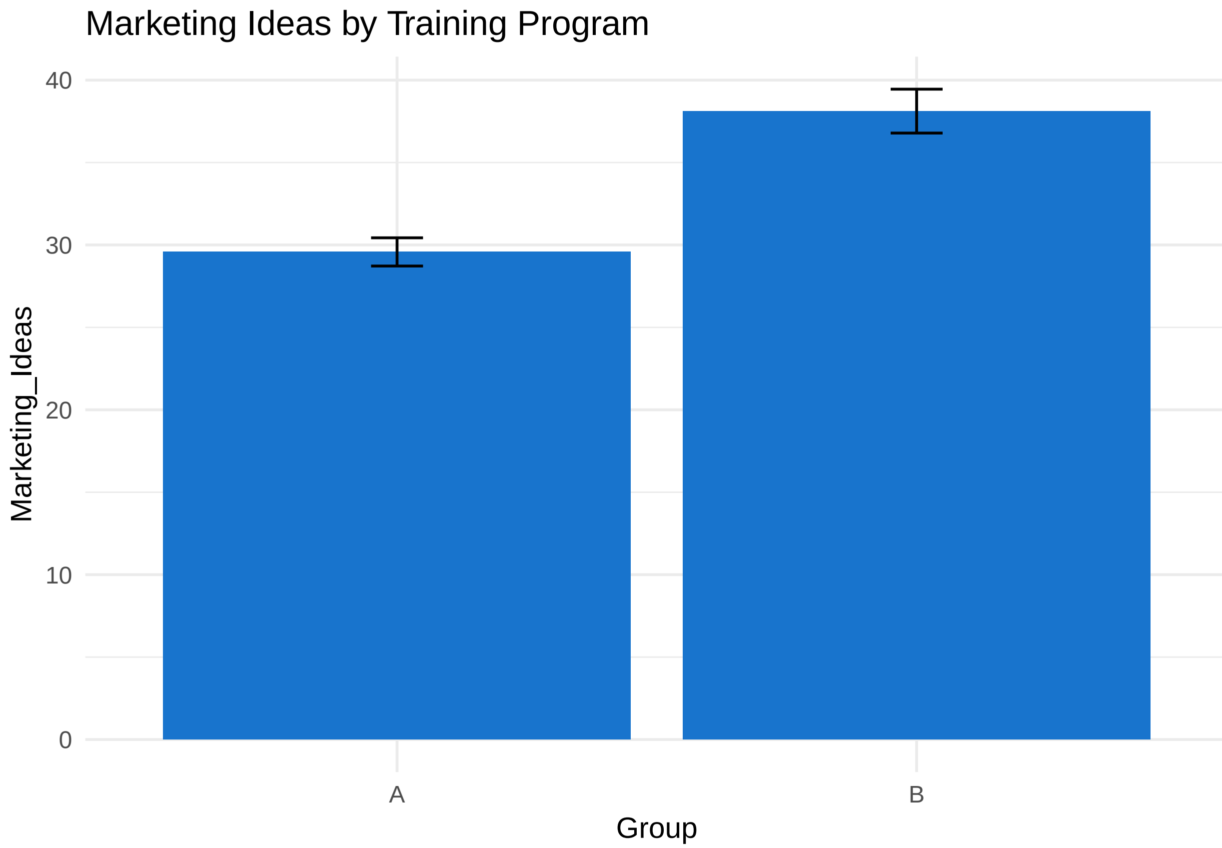
GGPLOT DATA

```
# Saving the summary data to call when putting in the ggplot2 code for the graph (for geom_errorbar).
sumdat <- summarySE(dat_A, measurevar="Marketing_Ideas", groupvars=c("Group"))
```

GGPLOT GRAPH

```
# creation of the bar graph - including specifications such as the color (#GoDodgers), title, addition
bar1 <- ggplot(sumdat, aes(x = Group, y = Marketing_Ideas)) +
  geom_bar(stat='identity', fill = 'dodgerblue3') +
  theme_minimal() +
  geom_errorbar(aes(ymin=Marketing_Ideas-se, ymax=Marketing_Ideas+se), width = .1)

bar1 + ggtitle('Marketing Ideas by Training Program')
```



Example 2: From your last study, you have some evidence that your week-long training works better than the one-day training for incoming marketing employees. You decide, this time, to see if the week-long training works significantly better for incoming marketing employees compared to those who have been with the program for a year. You don't have many incoming marketing employees who weren't in the first study, though. 15 of the incoming employees were not in the first study (Group A). You get 15 marketing employees that have been with the company for a year to participate as well (Group B). Everyone is placed in your new week-long training program. You are still collecting data on the amount of marketing ideas produced after the first month of being in the company. Based on your findings, does the training work better for incoming employees compared to employees that have been there for a year? Why?

Answer

> It is impossible to determine if the effect is due to training or not, as a baseline would need to be taken from the employees that have been with the company for a year to determine if their increase in knowledge was due to the training or not. However, based on the data analysis between incoming (Group A) and existing employees (Group B), after adjusting for failing homogeneity of variance assumption, there is no difference between groups A and B with number of marketing ideas one month after training.

IMPORT DATA

- View(dat_B)

```
dat_B <- read.csv("https://www.dropbox.com/s/qd4kcorra9js23i/independent.B.csv?dl=1")
```

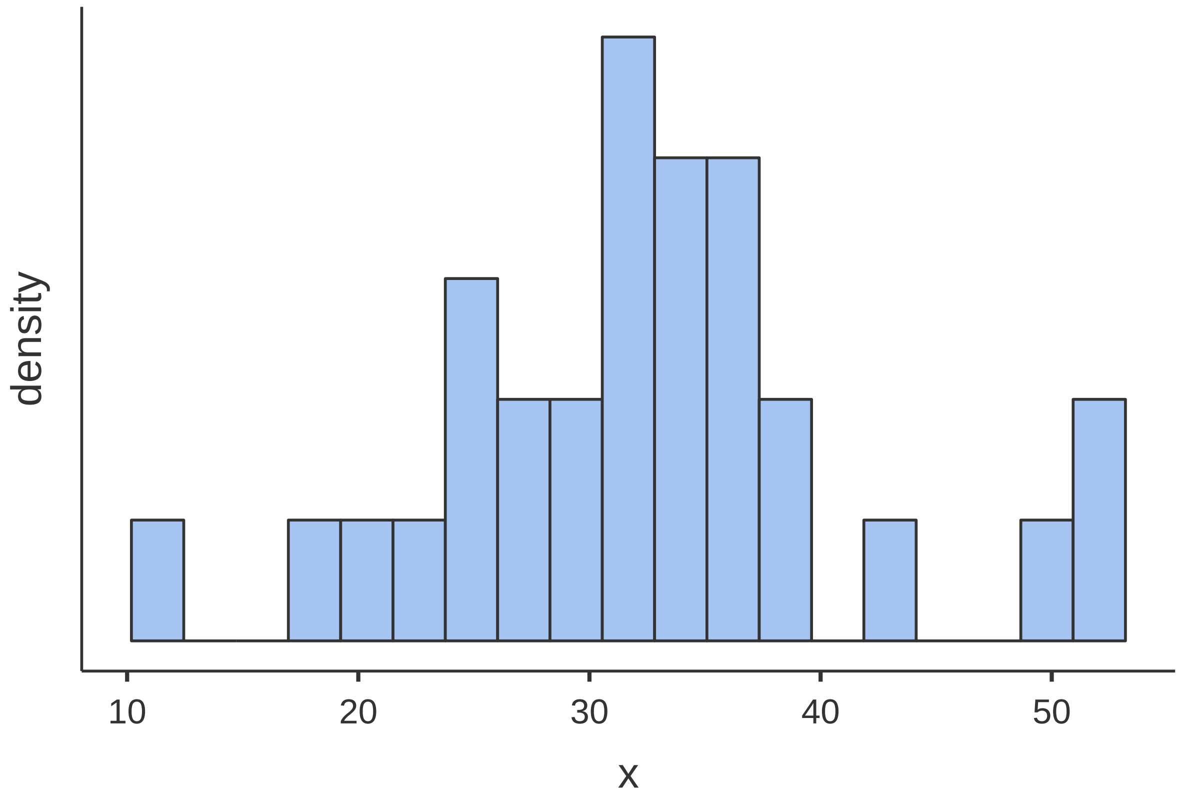
DESCRIPTIVES

Not split by group

We need it to be by group. Here the descriptives lump all the data in the data frame together * class(descB) —> What type of object is this? * class(dat_B) —> data.frame

```
descB <- descriptives(dat_B, vars = ('Marketing_Ideas'), hist = TRUE, sd = TRUE, se = TRUE, skew = TRUE
descB
```

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  ------------------------------------------
##                          Marketing_Ideas
##  ------------------------------------------
##    N                                  30
##    Missing                             0
##    Mean                             32.5
##    Std. error mean                  1.64
##    Median                           31.4
##    Standard deviation               9.00
##    Minimum                          11.2
##    Maximum                          51.9
##    Skewness                        0.199
##    Std. error skewness             0.427
##    Kurtosis                        0.718
##    Std. error kurtosis             0.833
##  ------------------------------------------
```



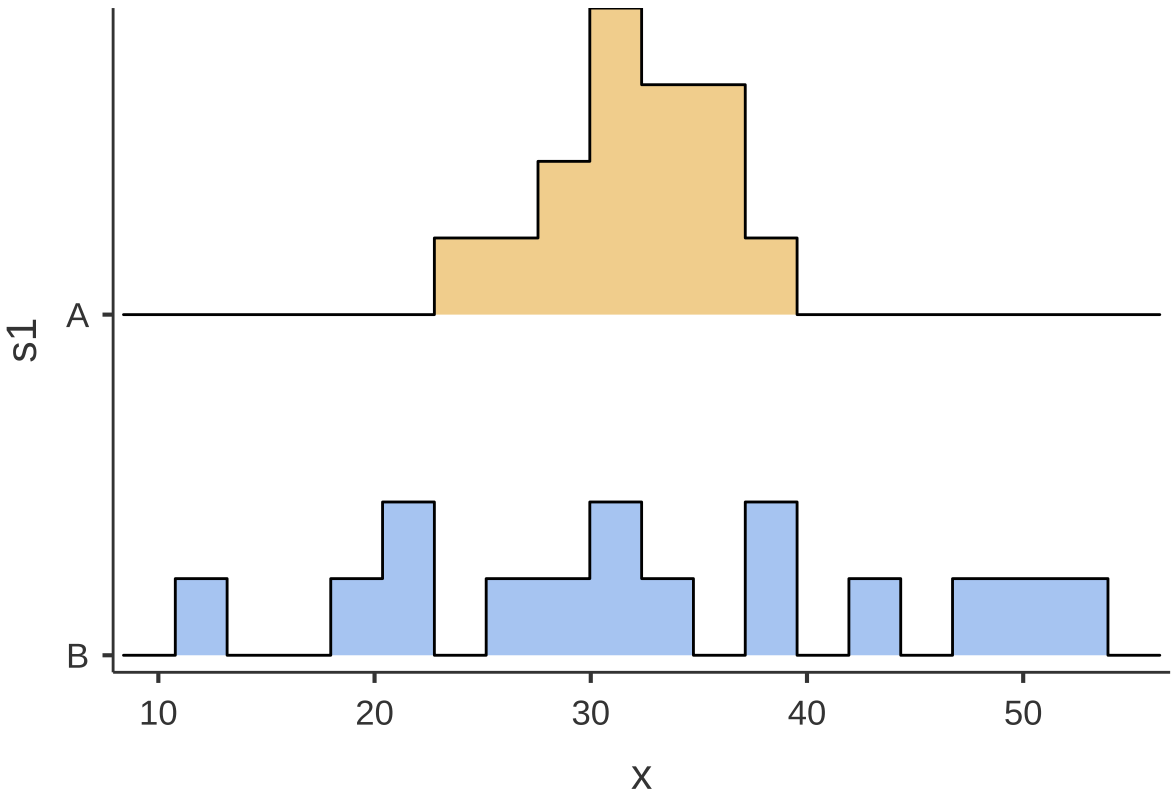
Split by group

We need this. * View(groupdescB)

```
groupdescB <- descriptives(dat_B, vars = ('Marketing_Ideas'), splitBy = 'Group', hist = TRUE, sd = TRUE
groupdescB
```

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  ---------------------------------------------------
##                           Group    Marketing_Ideas
##  ---------------------------------------------------
##    N                      A                     15
##                           B                     15
##    Missing                A                      0
##                           B                      0
##    Mean                   A                   31.9
##                           B                   33.2
##    Std. error mean        A                   1.04
##                           B                   3.17
##    Median                 A                   31.4
##                           B                   31.4
##    Standard deviation     A                   4.01
##                           B                   12.3
```

```
##    Minimum                 A                    24.8
##                            B                    11.2
##    Maximum                 A                    37.3
##                            B                    51.9
##    Skewness                A                  -0.349
##                            B                  0.0251
##    Std. error skewness     A                   0.580
##                            B                   0.580
##    Kurtosis                A                  -0.895
##                            B                  -0.810
##    Std. error kurtosis     A                    1.12
##                            B                    1.12
##  ---------------------------------------------------
```




```
dat_B$Group <- as.factor(dat_B$Group)
```

ANALYSIS

independent samples t-test

{jmv library} * vars = DV, group = IV, eqv = homogeneity of variance, welchs = welch's correction * we need to apply welch's correction after we find that levene's test fails (homogeneity of variance)

```
ttestIS(data = dat_B, vars = 'Marketing_Ideas', group = 'Group', eqv = TRUE, effectSize = TRUE, ci = TRU
```

```
##
##  INDEPENDENT SAMPLES T-TEST
##
##  Independent Samples T-Test
##  -----------------------------------------------------------------------------------------------
##                                      statistic    df       p        Lower    Upper    Cohen's d
##  -----------------------------------------------------------------------------------------------
##    Marketing_Ideas    Student's t     -0.384 <U+1D43>    28.0    0.704    -8.12     5.55       -0.14
##  -----------------------------------------------------------------------------------------------
##    <U+1D43> Levene's test is significant (p < .05), suggesting a violation of the assumption of
##    equal variances
##
##
##  ASSUMPTIONS
##
##  Test of Equality of Variances (Levene's)
##  ------------------------------------------
##                      F       df    p
##  ------------------------------------------
##    Marketing_Ideas    13.4     1    0.001
##  ------------------------------------------
##    Note. A low p-value suggests a
##    violation of the assumption of equal
##    variances
##
##
##  Group Descriptives
##  --------------------------------------------------------------------
##                      Group    N     Mean    Median    SD      SE
##  --------------------------------------------------------------------
##    Marketing_Ideas    A        15    31.9      31.4    4.01    1.04
##                       B        15    33.2      31.4    12.3    3.17
##  --------------------------------------------------------------------
```

```
# With Welchs
ttestIS(data = dat_B, vars = 'Marketing_Ideas', group = 'Group', welchs = TRUE, effectSize = TRUE, ci =
```

```
##
##  INDEPENDENT SAMPLES T-TEST
##
##  Independent Samples T-Test
##  -----------------------------------------------------------------------------------------------
##                                      statistic    df       p        Lower    Upper    Cohen's d
##  -----------------------------------------------------------------------------------------------
##    Marketing_Ideas    Student's t       -0.384    28.0    0.704    -8.12     5.55       -0.140
##                       Welch's t         -0.384    17.0    0.706    -8.32     5.76       -0.140
##  -----------------------------------------------------------------------------------------------
##
##
##  Group Descriptives
##  --------------------------------------------------------------------
##                      Group    N     Mean    Median    SD      SE
##  --------------------------------------------------------------------
##    Marketing_Ideas    A        15    31.9      31.4    4.01    1.04
##                       B        15    33.2      31.4    12.3    3.17
```

```
## ----------------------------------------------------------------------
```

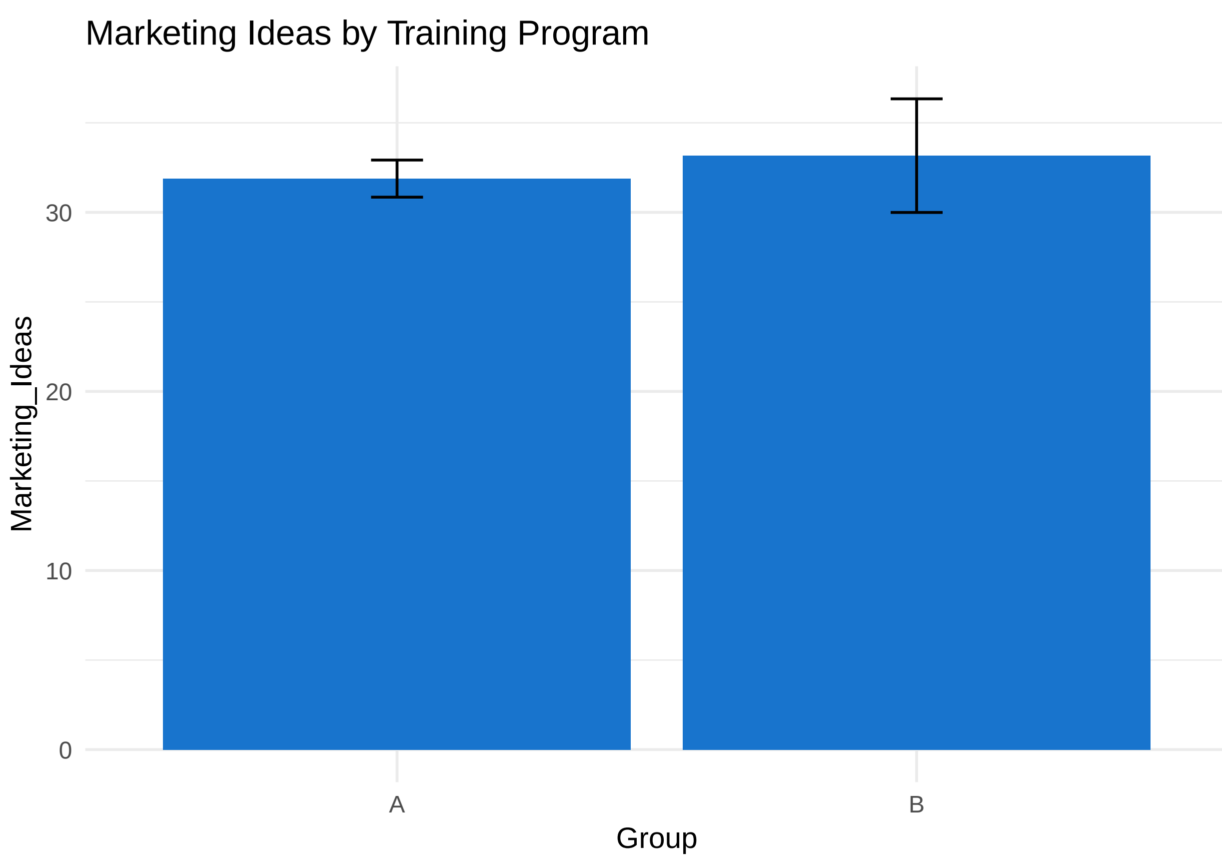
GGPLOT DATA

```
# Saving the summary data to call when putting in the ggplot2 code for the graph (for geom_errorbar).
sumdat2 <- summarySE(dat_B, measurevar="Marketing_Ideas", groupvars=c("Group"))
```

GGPLOT GRAPH

```
# Creation of the bar graph - including specifications such as the color (#GoDodgers), title, addition
bar2 <- ggplot(sumdat2, aes(x = Group, y = Marketing_Ideas)) +
  geom_bar(stat='identity', fill = 'dodgerblue3') +
  theme_minimal() +
  geom_errorbar(aes(ymin=Marketing_Ideas-se, ymax=Marketing_Ideas+se), width = .1)

bar2 + ggtitle('Marketing Ideas by Training Program')
```



Example 3: You want to see now if the training program is more effective for marketing different types of products. You get 40 employees who have not yet been through the training. Half of them and randomly assigned to creating marketing strategies for dog toys (Group A). The other half is randomly assigned to creating marketing strategies for children's books (Group B). You are still collecting data on the amount of marketing ideas produced after the first month of being in the company. Based on your findings, is there a difference for the training based on the type of product individuals are marketing? Why?

Answer

> The training for marketing strategies for dog toys group was (Group A) determined to have a more significant effect than the group for children's books. The reason is inconclusive, however, it may be due the differences in the training or within groups differences, however standard error was not significantly different in group A versus group B, which would account for quantitative difference in within groups variance.

IMPORT DATA

- View(dat_C)

```
dat_C <- read.csv("https://www.dropbox.com/s/nn8evxc5dhe3oyb/independent.C.csv?dl=1")
```

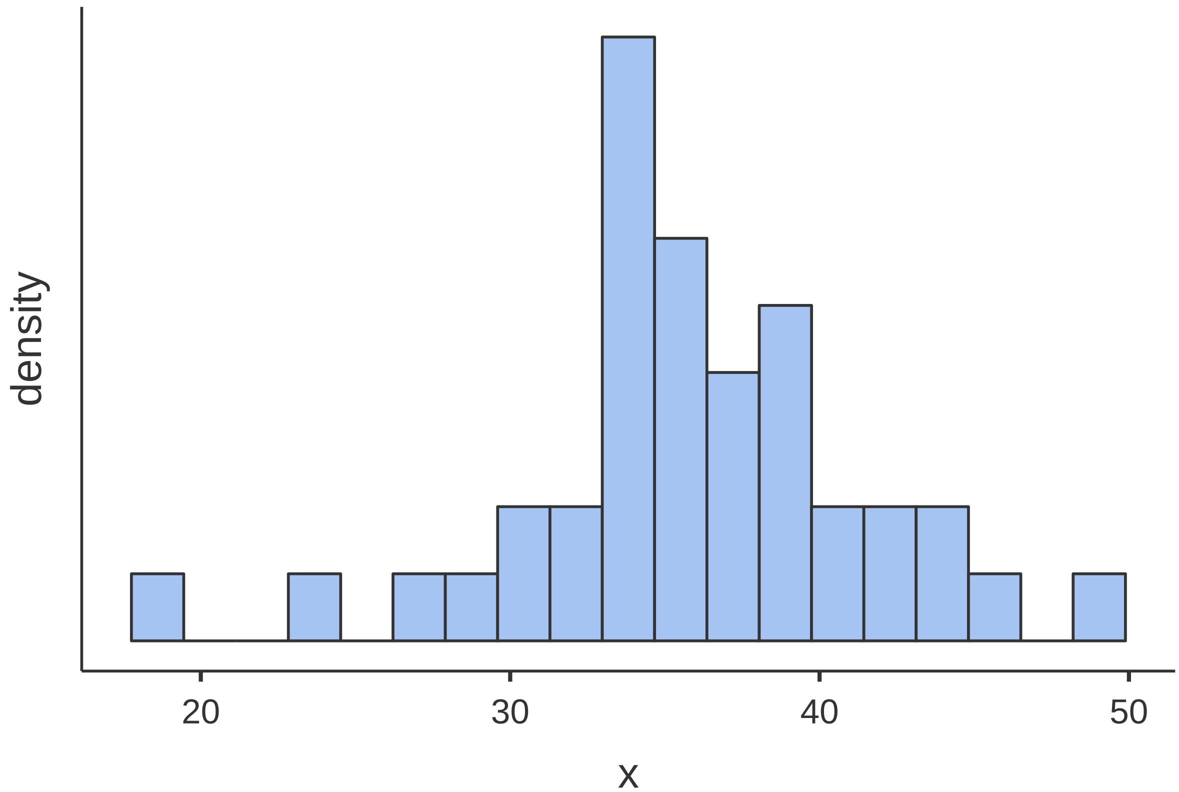
DESCRIPTIVES

Not split by group

We need it to be by group. Here the descriptives lump all the data in the data frame together * class(descC) —> What type of object is this? * class(dat_C) —> data.frame

```
descC <- descriptives(dat_C, vars = c('Marketing_Ideas'), hist = TRUE, sd = TRUE, se = TRUE, skew = TRUE
descC
```

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  -------------------------------------------
##                          Marketing_Ideas
##  -------------------------------------------
##    N                                  40
##    Missing                             0
##    Mean                             35.8
##    Std. error mean                 0.896
##    Median                           35.7
##    Standard deviation               5.67
##    Minimum                          18.2
##    Maximum                          48.7
##    Skewness                       -0.528
##    Std. error skewness             0.374
##    Kurtosis                         1.76
##    Std. error kurtosis             0.733
##  -------------------------------------------
```



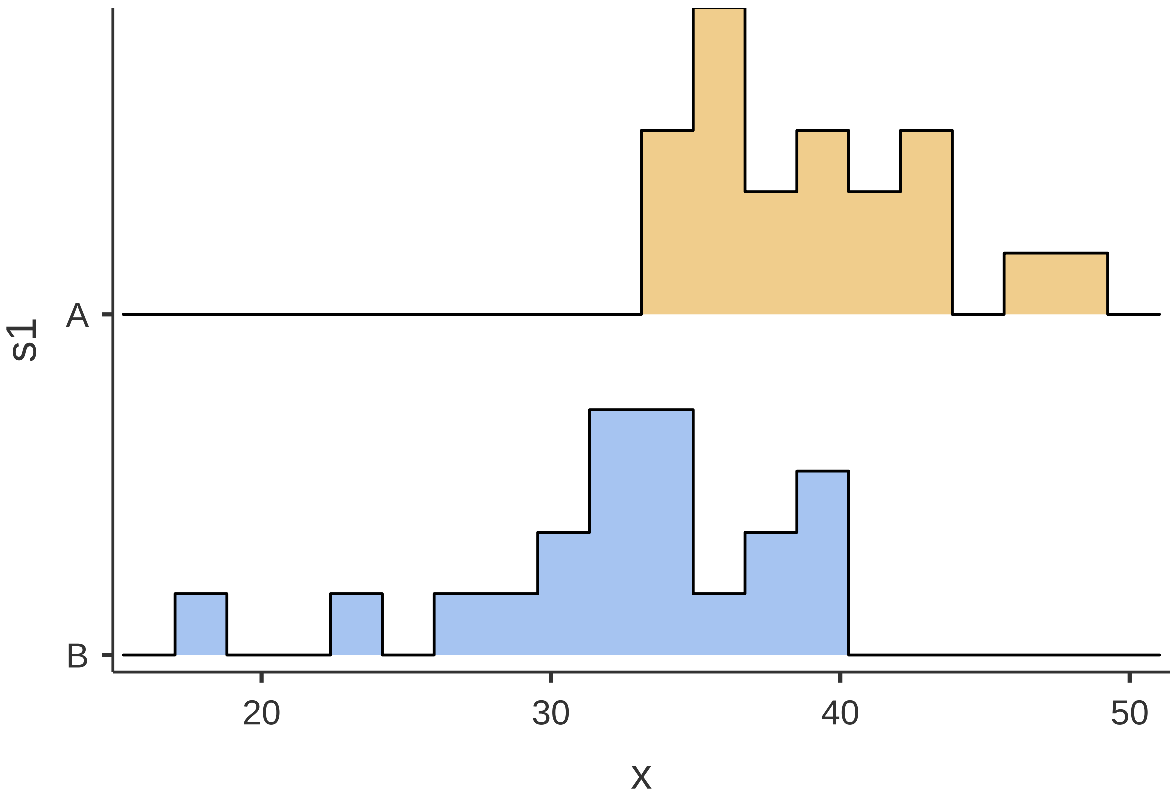
Split by group

We need this. * View(groupdescB)

```
groupdescC <- descriptives(dat_C, vars = c('Marketing_Ideas'), splitBy = 'Group', hist = TRUE, sd = TRUE
groupdescC
```

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  ---------------------------------------------------
##                           Group    Marketing_Ideas
##  ---------------------------------------------------
##    N                      A                     20
##                           B                     20
##    Missing                A                      0
##                           B                      0
##    Mean                   A                   38.9
##                           B                   32.7
##    Std. error mean        A                  0.949
##                           B                   1.18
##    Median                 A                   38.0
##                           B                   33.1
##    Standard deviation     A                   4.24
##                           B                   5.28
```

```
##    Minimum               A                     33.5
##                          B                     18.2
##    Maximum               A                     48.7
##                          B                     39.7
##    Skewness              A                    0.744
##                          B                    -1.11
##    Std. error skewness   A                    0.512
##                          B                    0.512
##    Kurtosis              A                   -0.160
##                          B                     1.81
##    Std. error kurtosis   A                    0.992
##                          B                    0.992
##  ---------------------------------------------------
```




```
dat_C$Group <- as.factor(dat_C$Group)
```

ANALYSIS

independent samples t-test

{jmv library} * vars = DV, group = IV, eqv = homogeneity of variance

```
ttestIS(data = dat_C, vars = 'Marketing_Ideas', group = 'Group', eqv = TRUE, effectSize = TRUE, ci = TRU
```

```
##
```

```
## INDEPENDENT SAMPLES T-TEST
##
## Independent Samples T-Test
## ------------------------------------------------------------------------------------------------
##                                     statistic    df      p         Lower    Upper    Cohen's d
## ------------------------------------------------------------------------------------------------
##   Marketing_Ideas    Student's t         4.07    38.0    < .001     3.10     9.23         1.29
## ------------------------------------------------------------------------------------------------
##
##
## ASSUMPTIONS
##
## Test of Equality of Variances (Levene's)
## -------------------------------------------
##                      F        df    p
## -------------------------------------------
##   Marketing_Ideas    0.133     1    0.718
## -------------------------------------------
##   Note. A low p-value suggests a
##   violation of the assumption of equal
##   variances
##
##
## Group Descriptives
## ------------------------------------------------------------------------
##                      Group    N     Mean    Median    SD      SE
## ------------------------------------------------------------------------
##   Marketing_Ideas    A        20    38.9      38.0    4.24    0.949
##                      B        20    32.7      33.1    5.28     1.18
## ------------------------------------------------------------------------
```

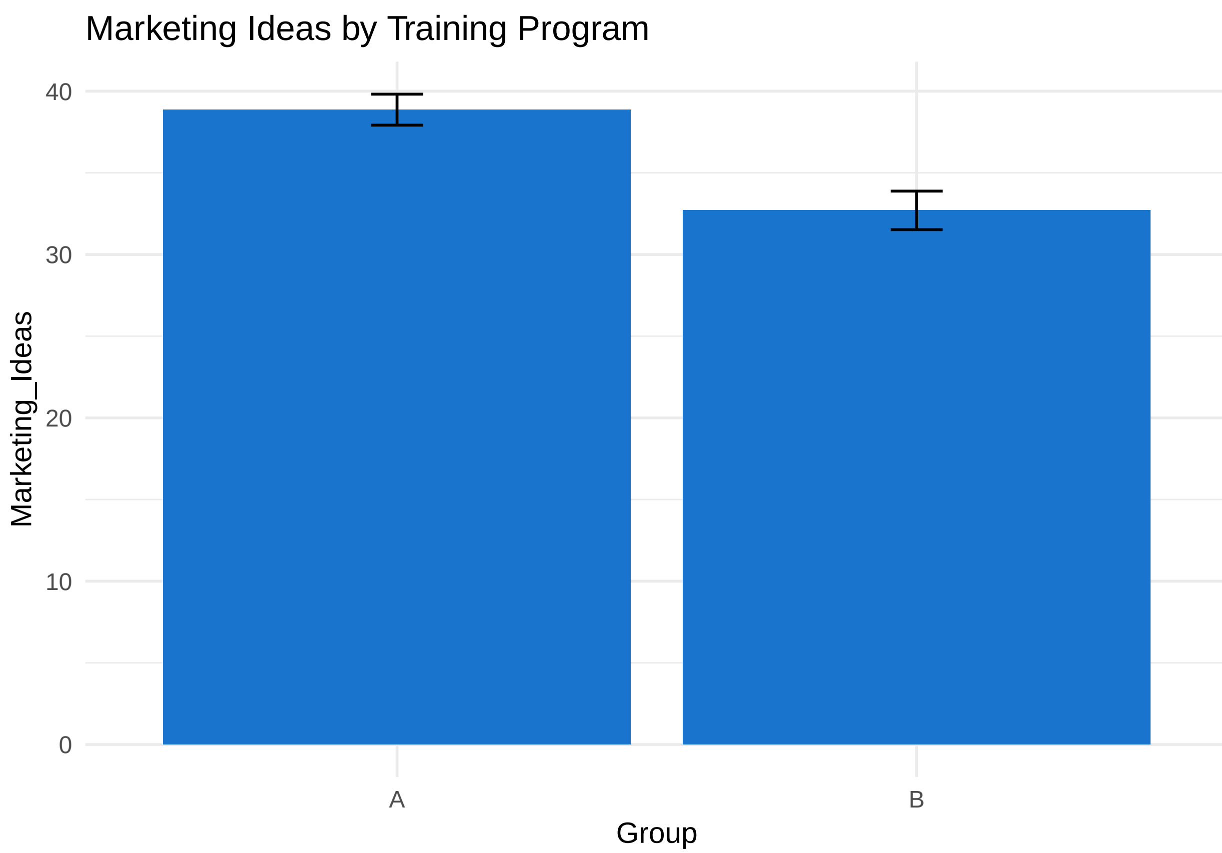
GGPLOT DATA

```r
# Saving the summary data to call when putting in the ggplot2 code for the graph (for geom_errorbar).
sumdat3 <- summarySE(dat_C, measurevar="Marketing_Ideas", groupvars=c("Group"))
```

GGPLOT GRAPH

```r
# creation of the bar graph - including specifications such as the color (#GoDodgers), title, addition
bar3 <- ggplot(sumdat3, aes(x = Group, y = Marketing_Ideas)) +
  geom_bar(stat='identity', fill = 'dodgerblue3') +
  theme_minimal() +
  geom_errorbar(aes(ymin=Marketing_Ideas-se, ymax=Marketing_Ideas+se), width = .1)

bar3 + ggtitle('Marketing Ideas by Training Program')
```


Marketing Ideas by Training Program
Marketing_Ideas
40
30
20
10
0
A
B
Group