# INVESTIGATING THE PEDAGOGY OF R VERSUS MINITAB IN TEACHING STATSTICS

**Reza O. Abbasian**
Texas Lutheran University
Department of Mathematics and Computer Science
Seguin, Texas, 78155 USA
rabbasian@tlu.edu

**John T. Sieben**
Texas Lutheran University
Department of Mathematics and Computer Science
Seguin, Texas, 78155 USA
jsieben@tlu.edu

## Abstract
In this paper we will investigate the advantages and shortcomings of the use of command line statistics software versus graphical user interface software in teaching elementary and advanced statistics topics. R will be used as an example of command line (CL) software and Minitab as an example of graphical user interface software (GUI).

**Keywords**: R, Minitab, basic statistics, Data analytics

## Introduction
Minitab has been in the market since the 1970's and has gone through major improvements since then, specifically the move from a command line software to a graphical interface about 25 years ago. Statistics software R has been around since year 2000. It is free and has become very popular among statisticians and data analytic experts. R is powerful but, many would argue, is not as user friendly as software with a GUI. Minitab is extremely user friendly and can handle basic statistical analysis at the undergraduate level and is considered an excellent "teaching" software for analysing small and medium size data, but unlike R, is not generally considered suitable for advanced statistical analysis involving big data. Texas Lutheran University is starting a new Master's program in Data analytics beginning in the fall 2017. The program calls for students to take two statistics courses beyond the introductory level. These two courses will be designed and taught by the authors as part of the Data Analytics core curriculum. The authors currently use Minitab in all our elementary statistics classes (statistics and applied statistics) but the new courses require a more powerful statistics software such as R which is the popular software among the practitioners in the fields of data analytics and data science. This is the motivation behind the current work. Our hypothesis is that Minitab is the preferred software for teaching basic concepts with small sets of data, whereas R is better suited for advanced topic with large data sets and that the "less friendly" code-base nature of R

helps students to develop a better understanding of statistical topics. In the following sections we will first introduce a summary of the comparison of the various aspects of the two packages, specifically: user control, ease of use, appearance of the output, multi-step analysis, targeted packages and the general capabilities when used in basic and advanced statistical analysis. We will continue our work by introducing a list containing the most common statistics commands and compare the syntax and ease of use between R and Minitab. We will then demonstrate the commands used to perform basic statistical tasks for several sample problems, using the same data set, focused on topics from basic to advanced statistics. We will briefly discuss the non-statistical usage of R, specifically using R to teach topics in calculus. Minitab assistant is the last topic of discussion in our paper.

**Comparing R and Minitab: A summary**
In the following table we compare R and Minitab in ten categories. Please note that the comparisons are made from the point of view of a novice student user. As the table indicates, R is clearly more suitable for multi-step tasks or projects in advanced statistics which involve using specific packages. An R user has more control over parameters to produce more meaningful output. Furthermore, R has better graphics capabilities and produces more appealing graphs. And finally (and most importantly!) R is free. Minitab, on the other hand, is much more user-friendly, has better tech support and more suited for elementary statistics projects.

Table 1: Comparing R and Minitab

| Attribute | Minitab | R |
|---|---|---|
| Cost | | x |
| Rigidity of Syntax | x | |
| Ease of Use | x | |
| Graphics | | x |
| Control | | x |
| Descriptive Stat | x | |
| Hyp Test | x | x |
| Advanced Stat | | x |
| Extendability | | x |
| Learning Curve | x | |
| Appearance | x | x |
| Tech Support | x | |
| Online Resources | x | x |
| Use in Other Disciplines | | x |
| Interface | | |

**Descriptive Statistics in R and Minitab**
Both systems will produce informative descriptive statistics with ease. In both systems you must first load a set of data on which to perform statistics. In Minitab your data is loaded into a spreadsheet like window, containing rows and columns in which your data resides. In R a typical data store would be a "data frame" which holds observations (rows) and variables being observed (columns). If operating R from the Integrated Development System (IDS) R Studio, a data viewer is available which will display a data frame in a manner that looks nearly identical to Minitab or Excel.

Minitab, because of its GUI interface requires a series of selections from drop down menus. For descriptive statistics on a variable, say weight, the sequence of menu selections will be STAT > Basic Statistics > Display Descriptive Statistics which will bring you to an interactive menu from which you select a list of variables for which you want descriptive statistics, a "by" variable which indicates how your data set should be segmented for the statistics (e.g. weight by gender), and finally a list of descriptive from which the user makes selections. In this example data on 54 anesthetized bears is taken from Mario Triola's companion website[i] to Elementary Statistics.

STAT > Basic Statistics > Display Descriptive Statistics

```
          Total
Variable  Count   Mean  Minimum    Q1  Median     Q3  Maximum
WEIGHT       54  182.9     26.0  84.5   150.0  242.5    514.0
```

R has separate commands for mean, median, standard deviation, etc., but the single command "summary(bears)" will give a report containing all the above for each numeric variable in the data set.

```
> summary(BEARS)
     AGE             MONTH            SEX           HEADLEN          HEADWTH           NECK            LENGTH
 Min.   :  8.00   Min.   : 4.000   Min.   :1.000   Min.   : 9.00   Min.   : 4.000   Min.   :10.00   Min.   :36.00
 1st Qu.: 17.00   1st Qu.: 8.000   1st Qu.:1.000   1st Qu.:11.50   1st Qu.: 5.000   1st Qu.:16.62   1st Qu.:50.50
 Median : 34.00   Median : 9.000   Median :1.000   Median :13.00   Median : 6.000   Median :20.00   Median :60.75
 Mean   : 43.52   Mean   : 8.407   Mean   :1.352   Mean   :12.95   Mean   : 6.194   Mean   :20.56   Mean   :58.62
 3rd Qu.: 57.75   3rd Qu.:10.000   3rd Qu.:2.000   3rd Qu.:14.50   3rd Qu.: 7.000   3rd Qu.:24.00   3rd Qu.:66.12
 Max.   :177.00   Max.   :11.000   Max.   :2.000   Max.   :17.00   Max.   :10.000   Max.   :31.50   Max.   :76.50
     CHEST           WEIGHT
 Min.   :19.00   Min.   : 26.0
 1st Qu.:29.00   1st Qu.: 87.0
 Median :34.00   Median :150.0
 Mean   :35.66   Mean   :182.9
 3rd Qu.:43.50   3rd Qu.:232.0
 Max.   :55.00   Max.   :514.0
```
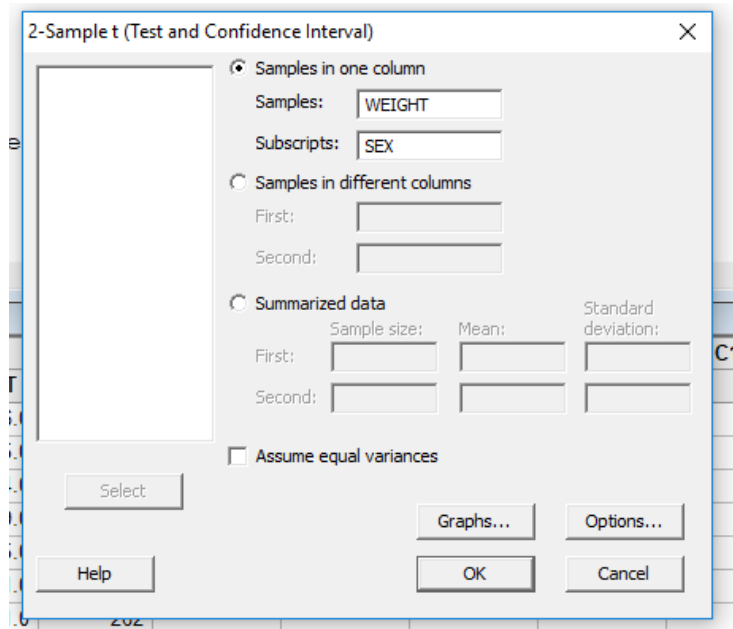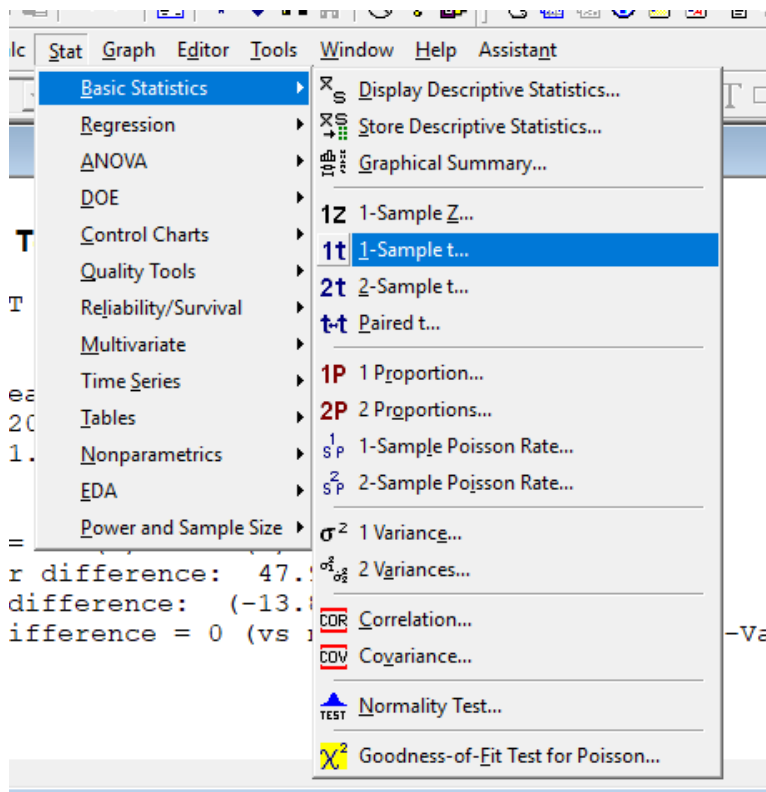
For descriptive statistics both systems seem about equally user friendly.

**Hypothesis testing in R and Minitab**
Is there a significant difference in the average weight of male and female bears? This question would be addressed by a hypothesis test, specifically with a two-sample t-test. Both Minitab and R use the Welch Two Sample t-test. The only difference is in how one summons the test. In Minitab the call is through a series of clicks on tabs or dropdown menus. STAT > Basic Statistics > 2-Sample-t which brings you to an interactive menu

where you identify the data for the test or supply summarized values (sample size, mean, s.d. for each variable).

lc  Stat  Graph  Editor  Tools  Window  Help  Assistant

Basic Statistics ▶
Regression ▶
ANOVA ▶
DOE ▶
Control Charts ▶
Quality Tools ▶
Reliability/Survival ▶
Multivariate ▶
Time Series ▶
Tables ▶
Nonparametrics ▶
EDA ▶
Power and Sample Size ▶

X̄s Display Descriptive Statistics...
X̄S Store Descriptive Statistics...
Graphical Summary...

1Z  1-Sample Z...
1t  1-Sample t...
2t  2-Sample t...
t-t  Paired t...

1P  1 Proportion...
2P  2 Proportions...
sP  1-Sample Poisson Rate...
sP  2-Sample Poisson Rate...

σ² 1 Variance...
σ²² 2 Variances...

COR  Correlation...
COV  Covariance...

TEST  Normality Test...
χ²  Goodness-of-Fit Test for Poisson...

r difference:  47.
difference:  (-13.
ifference = 0 (vs

**2-Sample t (Test and Confidence Interval)**           ✕

⦿ Samples in one column
  Samples:     WEIGHT
  Subscripts:  SEX

○ Samples in different columns
  First:
  Second:

○ Summarized data
                Sample size:   Mean:   Standard deviation:
  First:
  Second:

☐ Assume equal variances

  Select

  Help        Graphs...   Options...
              OK          Cancel

Minitab returns a nicely formatted output with summary data for the two samples, the value of the t-statistic, the p-value, and a confidence interval estimate for the difference of the two population means.

```
Two-Sample T-Test and CI: WEIGHT, SEX

Two-sample T for WEIGHT

                         SE
SEX   N   Mean   StDev  Mean
1    35    200    134    23
2    19  151.8   90.4    21


Difference = mu (1) - mu (2)
Estimate for difference:  47.9
95% CI for difference:  (-13.8, 109.6)
T-Test of difference = 0 (vs not =): T-Value = 1.56  P-Value = 0.125  DF = 49

|
```

Using R to complete the same test involves entering one command with appropriate parameters:

```
> t.test(BEARS$WEIGHT~BEARS$SEX,BEARS,var.equal = FALSE)

        Welch Two Sample t-test

data:  BEARS$WEIGHT by BEARS$SEX
t = 1.5596, df = 49.38, p-value = 0.1252
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -13.80953 109.61104
sample estimates:
mean in group 1 mean in group 2
       199.7429        151.8421
```

The R input is more compact than walking through a series of menus, but it does involve remembering some syntax. The parameters (in order) inform R that we wish to test equality of means for bears partitioned by sex, the data set is called "bears," and it is not assumed that the variances are equal. From the R help files we can instantly view the full set of parameters and their default values for the Welch Two Sample t-test.

```
t.test (x, y = NULL,
        alternative = c("two.sided", "less", "greater"),
        mu = 0, paired = FALSE, var.equal = FALSE,
        conf.level = 0.95, ...)
```

The above serves the purpose of a GUI menu but only appears if you request it. It tells us, among other things, that t.-test will perform a one sample t-test if we supply a single variable instead of two (the second variable "y" has a default value of "NULL"), and independent samples with no assumption of equal variances are default values.

## Graphics in R and Minitab

The graphics in Minitab are easy and give good results, but the graphics in R are spectacular. R recognizes 657 color names, that's an invitation to get decorative with your graphs.
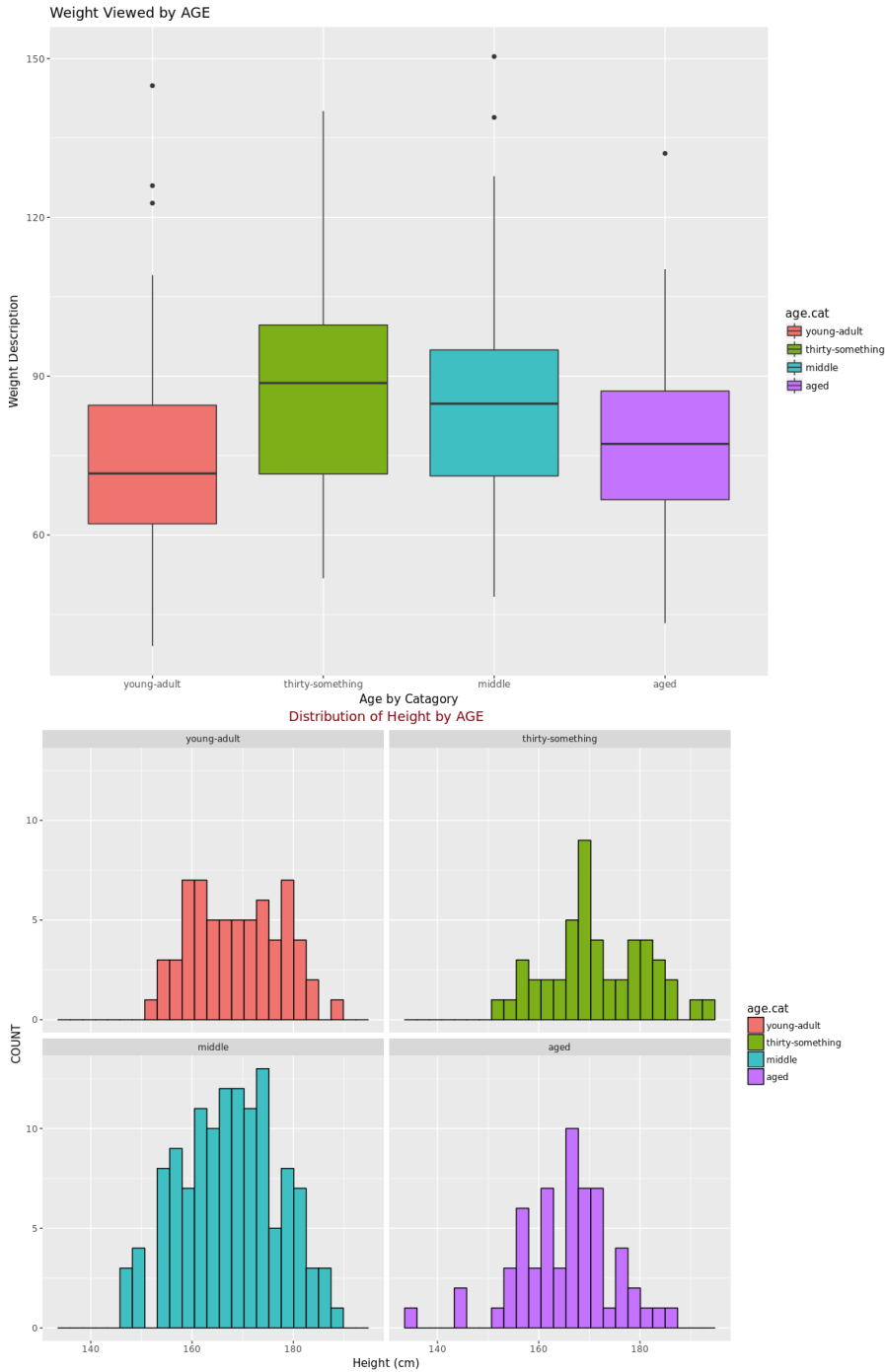


Figure1: Graphics in R

As remarked above, Minitab graphs are easy to create; default values will be adequate for many situations. The graphs are customizable but that is not quite as transparent. For example, to change colors and bin sizes on the following one must start by a right click in one of the cells.
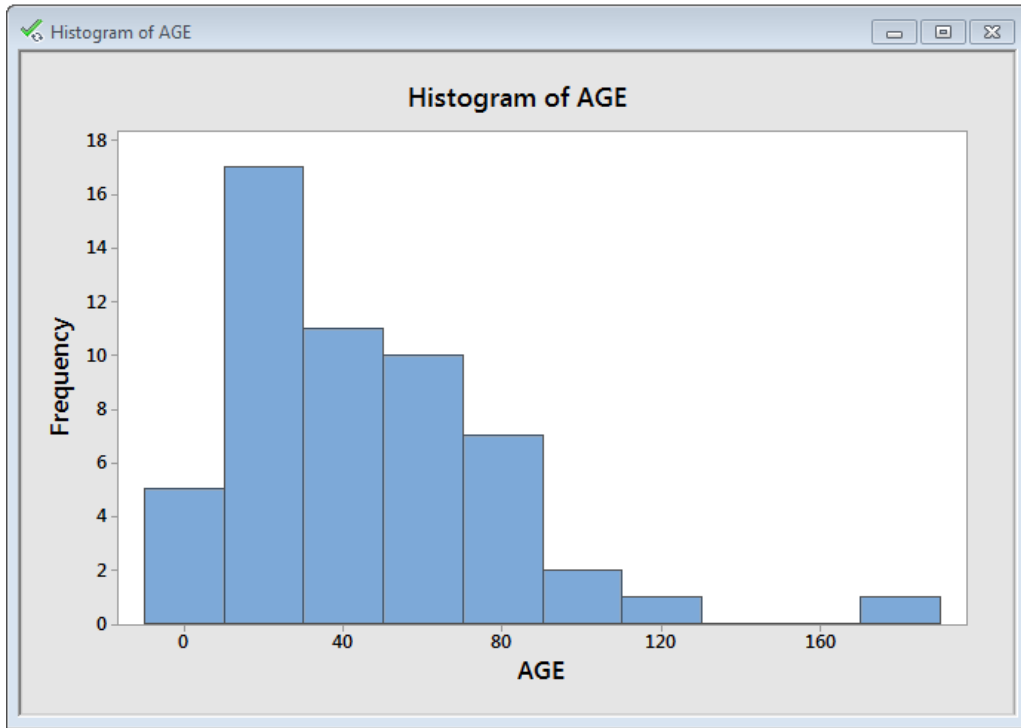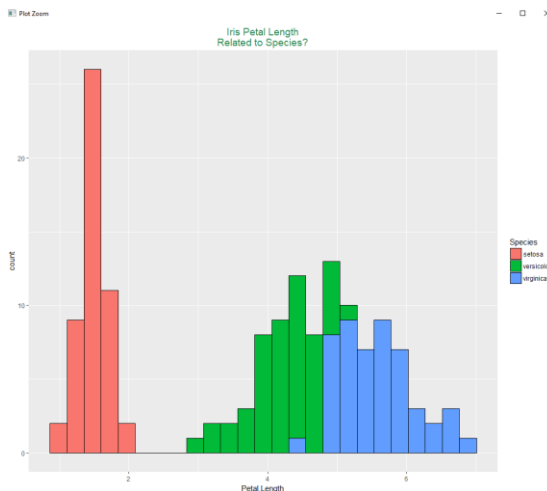


Figure 2: Graphics in Minitab

Right click in a bin is not an intuitive move, but the syntax for creating graphs in R is also far from intuitive.

```
gf_histogram(~Petal.Length,data = iris,color="black",fill=~Species)%>%
gf_labs(title="Iris Petal Length\nRelated to Species?")+
theme(plot.title = element_text(hjust = .5,vjust = 0,color="seagreen",size=14))
```



Within a given graphics package, and there are several, a predictable order of parameters makes creating graphs easier, but practice is required to make the process feel natural. That is perhaps just the nature of the command line interface.

Figure 3: Graphics in R

In Minitab it is easy to create a scatter diagram with one parameter, for example: bear weight predicted by length with color coding for gender or weight predicted by length with the shape of the mark indicating age.
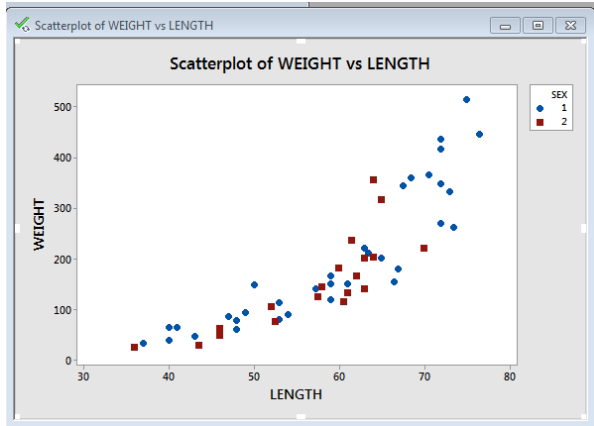


Figure 4: Scatter diagram in Minitab

In R these can easily be combined into a single scatter plot in which position indicates length and weight, color of dot indicates gender and size of dot indicates age.
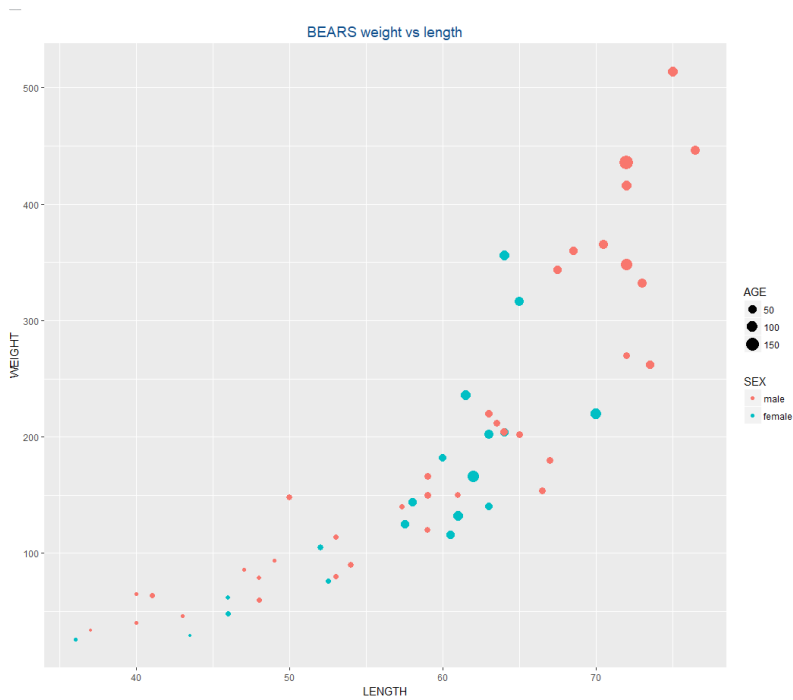


Figure 5:  Scatter diagram in R

## Advanced Statistics in R and Minitab

Admittedly the term "advanced statistics" is a bit ambiguous, but from the perspective of the novice user of statistical software we feel that "advanced topics" include piecewise regression, logit regression, non-polynomial regression, forecasting, nonparametric statistics, and programming. As a first illustration of these topics we will look at piecewise linear regression.
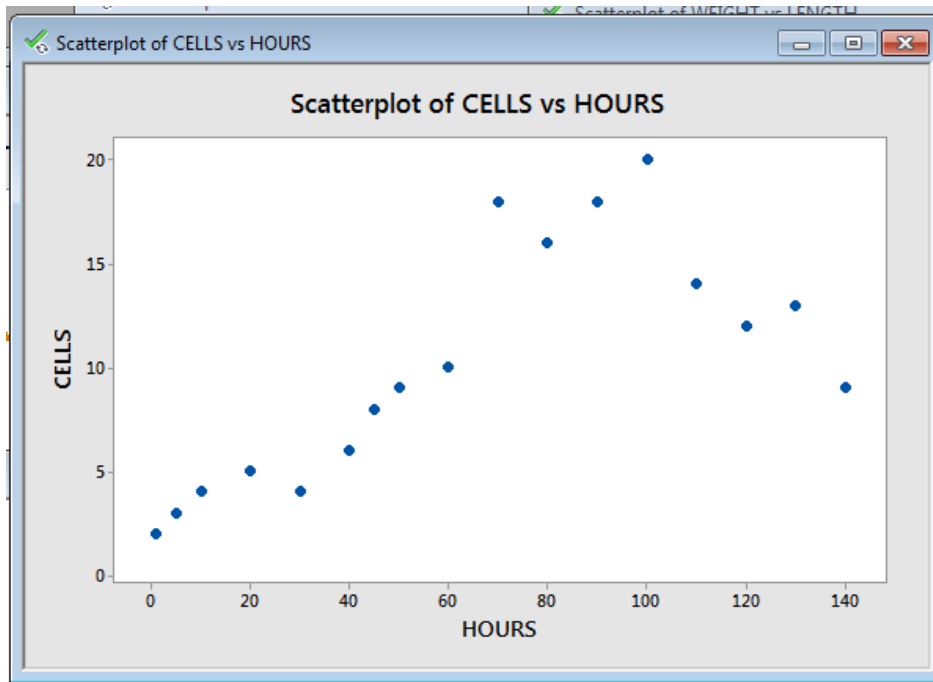


Figure 6: Scatter diagram of a piecewise linear pattern

This graph represents number of cells growing as a function of time. On the left there appears to be a linear relation with a positive slope but somewhere around 100 hours the slope appears to abruptly change to negative. This data set may be modeled by two different linear functions, depending on the size of the time variable. Minitab is able to do piecewise linear regression with the user supplying a value for the point where the change in slope occurs which is called a "break point" or a "knot."

## Results for: GROWTH.MTP

## Code

## Summary

| Lower End | Upper End | Recoded Value | Number of Rows |
|---|---|---|---|
| 0 | 70 | 0 | 10 |
| 70 | 140 | 1 | 7 |

```
Regression Equation

CELLS = 0.17 + 0.2221 HOURS - 0.2729 (Hours - 70) * U

U = 0 if Hours ≤ 70, U = 1 if Hours > 70
Fits and Diagnostics for Unusual Observations
```
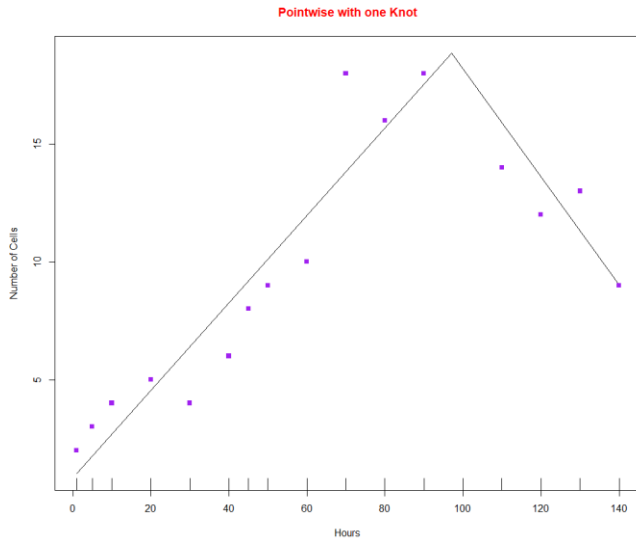
R will also perform piecewise regression. R requires an initial guess at the location of the knot but will optimize the position of the knot. To do the regression, produce a graph, and label axes and the graph itself does require some syntax.

```
try2.seg <- segmented(try1.lm, seg.Z = ~ HOURS,
psi=list(HOURS = 70))

plot(try2.seg, main = "", xlab="", ylab="") +
points(GROWTH$CELLS~GROWTH$HOURS, pch =
15, col = "purple", bg="gold")+


title(main = "Pointwise with one Knot", col.main
="red", font = 3,
      xlab = "Hours", ylab = "Number of Cells")
```
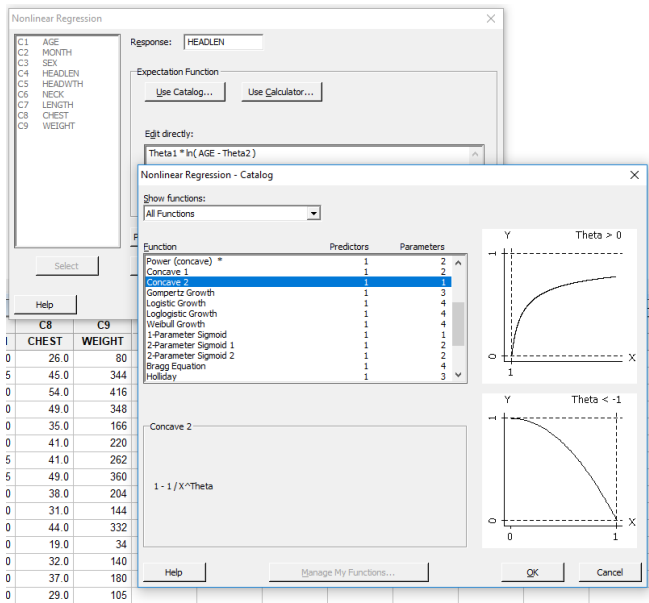
The graphic at the left is created with base graphics, that is, graphics native to R and not requiring the loading of special purpose code in the form of "packages."

R will easily solve similar problems with multiple knots.

Figure 7:  Piecewise linear model in R



For nonlinear regression Minitab has a well-developed interface that gives the user a wide selection of models from which to choose.

The nonlinear regression models are available in R also, though they are not quite as easy to create.

Figure 8: Nonlinear Regression in Minitab

**Use of R in other disciplines**
R is most often described as "a free software environment for statistical computing and graphics." None the less, we have heard of colleagues who are experimenting with R as a language for use in Calculus and Differential Equations classes. We are not involved in these experiments but are curious and await reports on their success.

**Minitab Assistant**
The Minitab assistant is a significant addition to the basic Minitab program. The goal is multifold. It offers step by step assistance with: picking the right tool, running your analysis, interpreting your results, and sharing your findings. All of this is done through a very attractive graphical user interface. For illustration two of the four assistant pages follow.
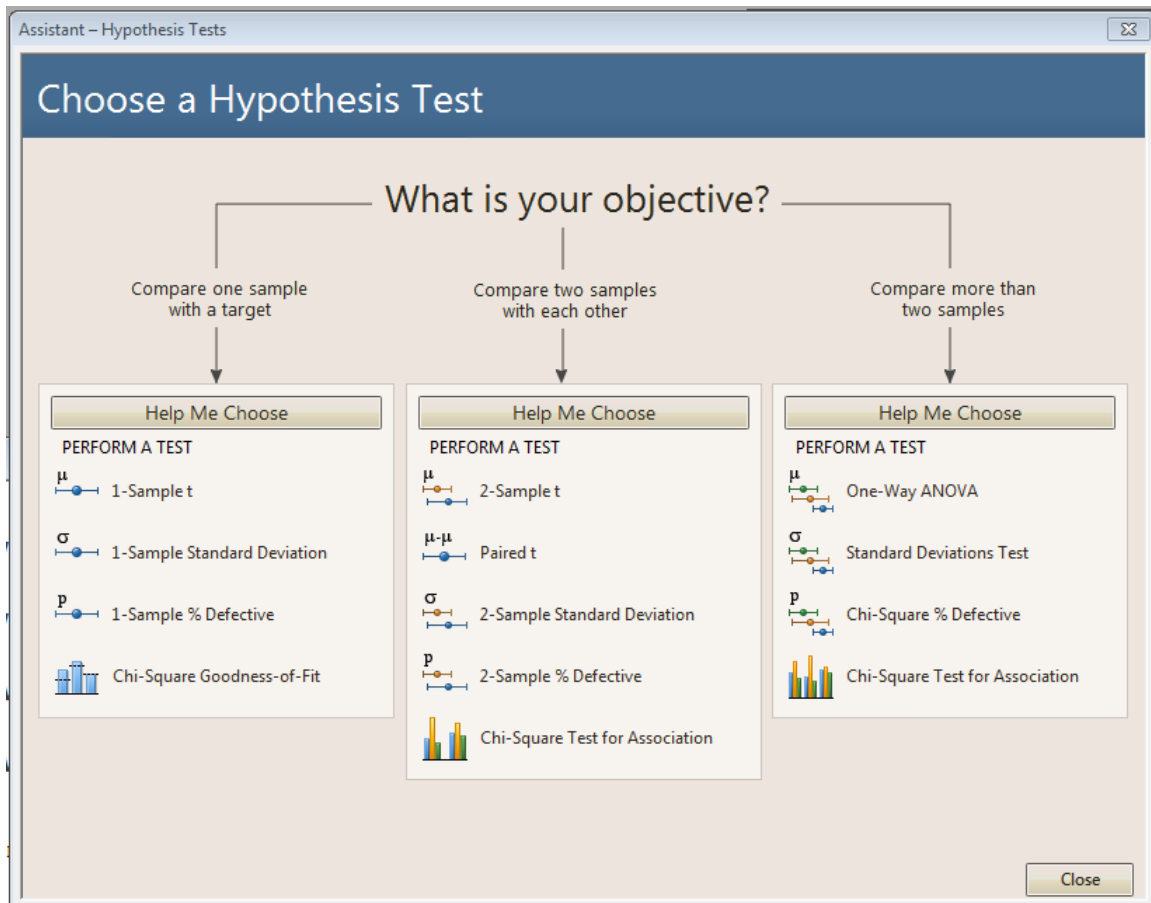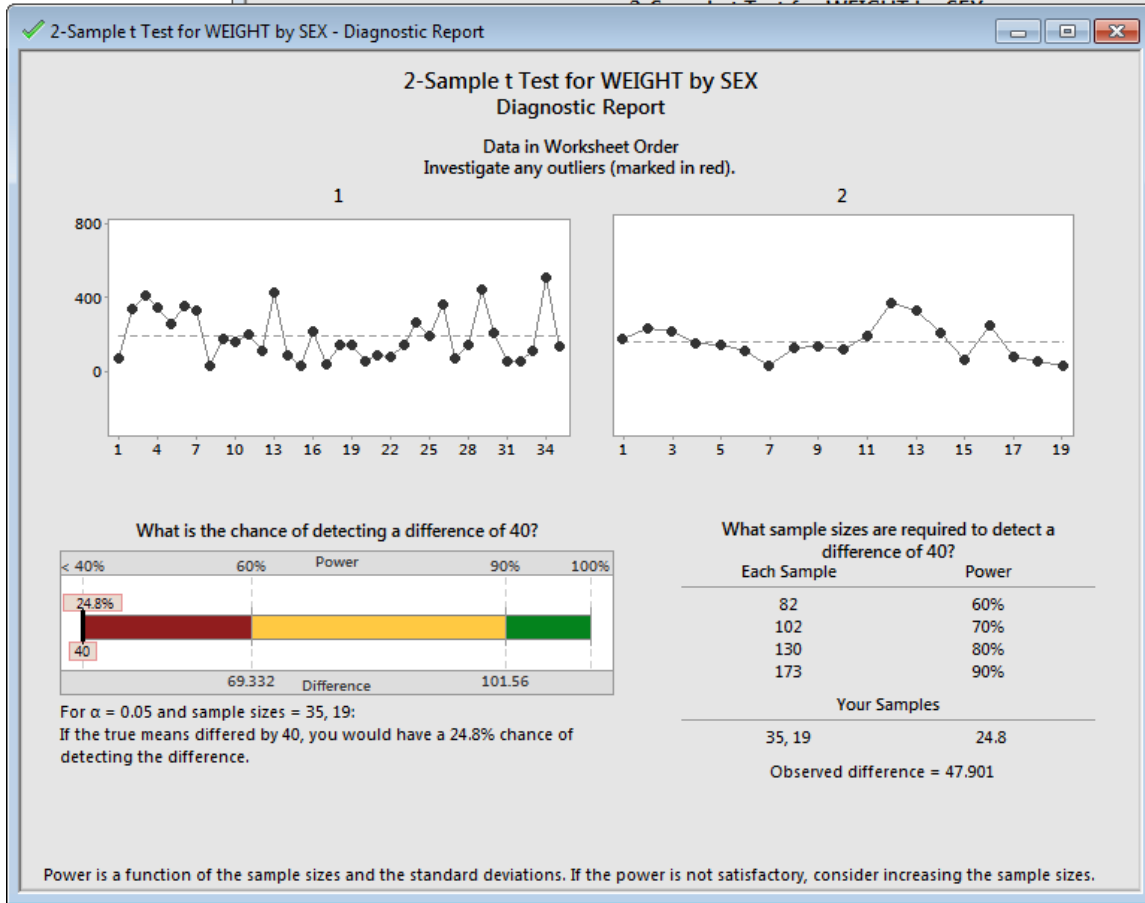


Figure 9: Minitab Assistant

Figure 10: Results from Minitab Assistant

**Conclusions**

Is the GUI interface better? When using R my students have been heard to complain, "I can't remember which buttons to push to get that result." Still the interface is designed to be a set of reminders or prompts for which step comes next and the user is not responsible for syntax. Minitab commands do not fail to work or give erroneous results because of a misplaced comma. R, as a result of its structure that embraces user defined "packages" to do special things and to refine the work of routine things, is ever growing and broadening its capabilities. But this ability to quickly respond to new demands comes at a price. The command line has rigid syntax and small errors can prevent execution. This is often quite frustrating to new users of R. The built- in help system of R (and indeed this is true of Minitab as well) offers immediate reminders of correct syntax but to the uninitiated the reminders and error messages can be cryptic and overwhelming.

As mentioned earlier, R is free and Minitab is a commercial product that must be licensed. The cost of Minitab is quite reasonable and should not present a barrier to its use by any university.

R, because of its growing reputation in the research community will have a certain name recognition and appeal to users, especially professionals venturing well beyond introductory applications. On a personal level we like both. At this time, we lean a bit toward Minitab for our first course in statistics, but appreciate the power and beauty of R for more advanced courses.

**References**
[1] Mendenhall, Sincich,  A Second Course in Statistics: Regression Analysis, 7th edition, Prentice Hall, 2012
[2]  Triola, Mario, Elementary Statistics, 12th edition, Pearson, 2014
[3] Cotton, Richard, Learning R, O'Reilly, 2013