

STAT CRAFT



WEB-BASED SOLUTION FOR

 **ANALYSIS WITH R**

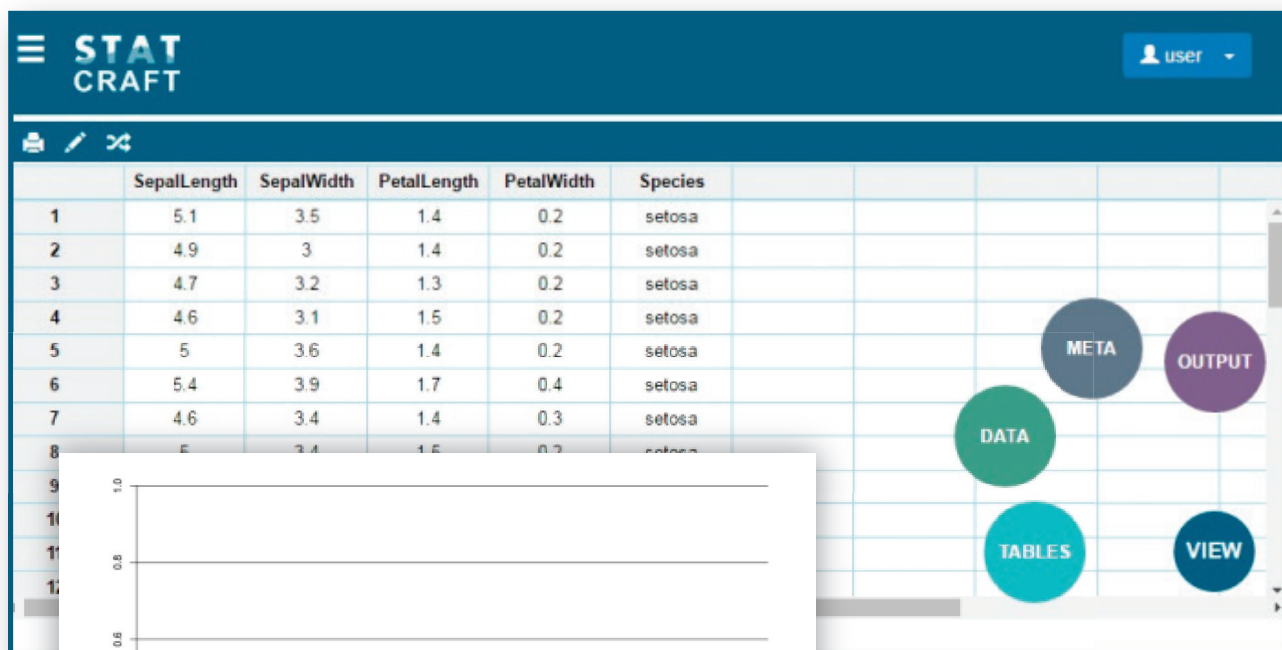
MINUS THE HASSLE OF CODING

What is STATCRAFT?

STATCRAFT is a browser based rich GUI that helps Data Scientists harness the power of R without having to write a single line of code. With STATCRAFT you can easily bring in and organize your data, access some of the most popular data analysis techniques in R and view the results in

Why STATCRAFT ?

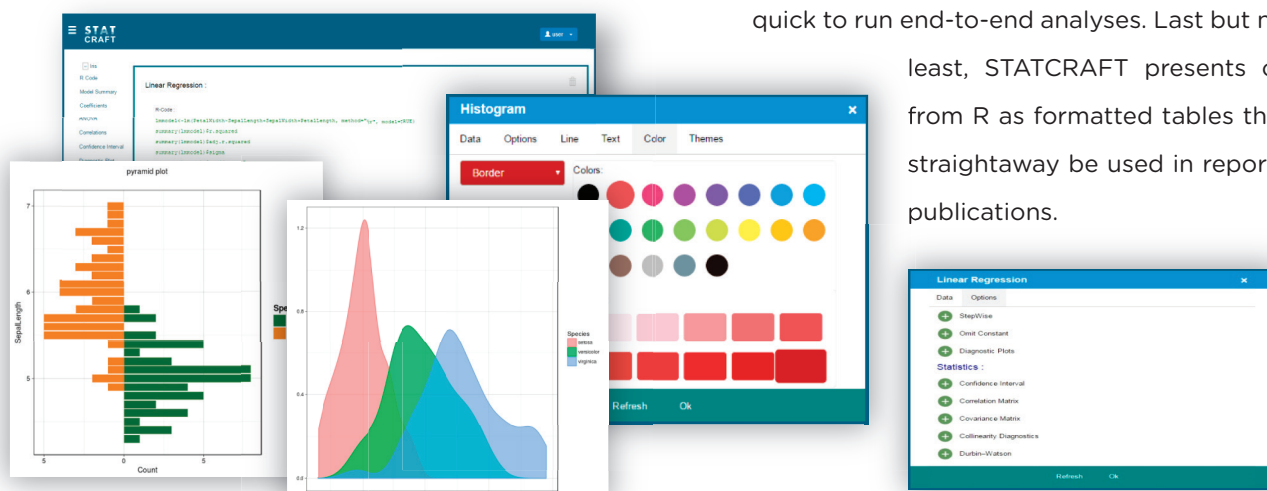
R is an extremely popular and powerful analytics software that is rapidly becoming the tool of choice for Data Scientists around the world. However, R does pose a few challenges. First, in R any analysis requires coding, consuming time and effort. With STATCRAFT running an analysis is as simple

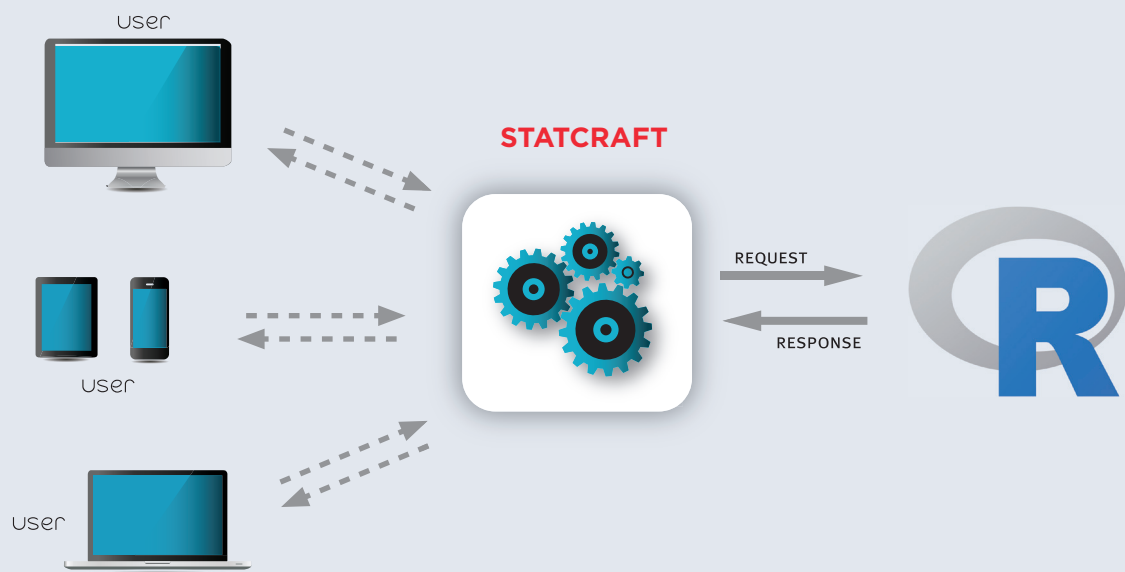


elegantly formatted output tables. STATCRAFT makes it simple and easy to focus on analysis rather than programming.

as choosing the options from the menu. Second, the wide number of packages available in R can often require the user to comb through a large number of resources to identify the set of techniques that best suit the task at hand. STATCRAFT combines and groups all the related techniques together making it easy and quick to run end-to-end analyses. Last but not the

least, STATCRAFT presents output from R as formatted tables that can straightaway be used in reports and publications.





The STATCRAFT Architecture

STATCRAFT runs on the organization's web server, allowing users to access R functions through their browsers. While the data stays secure on the server, the users have the comfort of working from the familiar environment of their PCs and devices.

The server based architecture eliminates the need for system administrators to manage multiple installations while making it easier to monitor usage and the organisations data resources from a single location.

SALIENT FEATURES

► Frequencies

► **Descriptives:** Count, Minimum, Maximum, Range, Mean, Sum, Standard Deviation, Variance, Standard Error of Mean, Coefficient of Variation, Kurtosis, Skewness

► **Charts:** Bar Plot, Histogram, Scatter Plot, Pie Chart, Box Plot

► **t-Test:** One Sample t-Test, Independent Samples t-Test, Paired Sample t-Test

► **One Way ANOVA:** Holm, Hochberg, Hommel, Bonferroni, BH, BY, fdr, None, Tukey's HSD Plot, Bartlett Test, ANOVA Plot, Levene's test

► **Two-Way ANOVA:** Holm, Hochberg, Hommel, Bonferroni, BH, none, BY, fdr, Tukey's HSD, Levene's Test, Tukey HSD Plot, Residuals vs fitted Plot, Normal Q-Q Plot, Scale-location Plot, Cook's distance Plot, Residual vs leverage Plot, Cook's distance vs leverage Plot, Interaction Plot

► Non Parametric Tests: *Kruskal-Wallis, Wilcoxon*

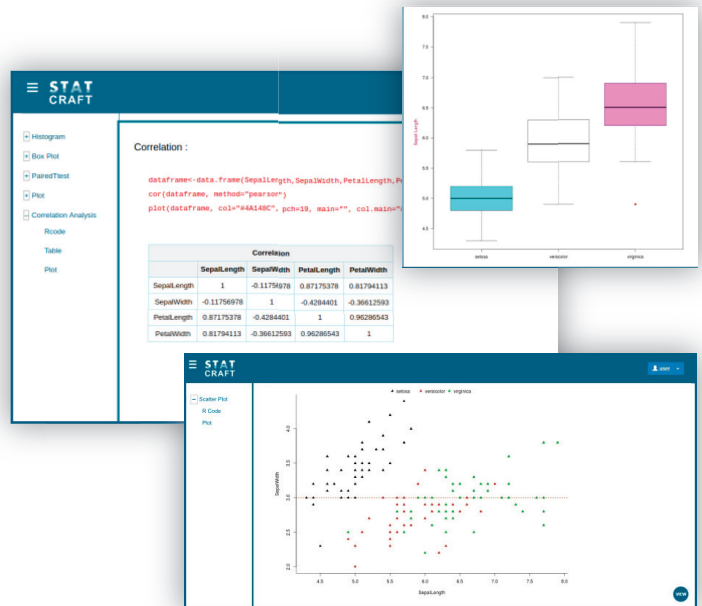
► **Crosstabs:** Count, Total, Row Proportion, Column Proportion, Table Proportion, Row Sum, Column Sum, Expected, Standardized Residuals, Adjusted Standardised Residuals, Residuals, Chi-Square Proportion, Chi-Square Statistics, Chi-Square Statistics Correlation, McNemar, Fisher

► **Group Descriptives:** Mean, Maximum, Minimum, Standard Deviation, Variance, Median

► **Correlation:** Pearson, Spearman, Kendall, SPLOM

► **Linear Regression:** Stepwise: Forward, Backward, Both. Confidence Intervals, Correlation Matrix, Covariance Matrix, Dubin-Watson, VIF, Residuals vs fitted Plot, Normal Q-Q Plot, Scale Location Plot, Cook's Distance Plot, Residuals vs Leverage Plot. Cook's distance vs Leverage

- ▶ **Binomial Logistic Regression:** Stepwise, ANOVA, Confidence Interval, Odds Ratio, covariance matrix, correlation matrix, Variable importance, Pseudo R square, Hosmer Lemeshow Goodness of fit, Likelihood Ratio Test, Wald Test, Classification Table, Classification Rate, ROC Curve.
- ▶ **Multinomial Logistic Regression:** Stepwise, ANOVA, Confidence Interval, Odds Ratio, Correlation Matrix, Variable Importance, Classification Table, Likelihood Ratio Test, Pseudo R2, Pearson chi-Square
- ▶ **K-Means Cluster Analysis:** Hartigan-Wong, Lloyd, Forgy, MacQueen
- ▶ **Hierarchical Cluster Analysis:** Distance - Euclidean, Maximum, Manhattan, Canberra, Binary, Minkowski. Method - ward.D, ward.D2, Single, Complete, Average, Mcquitty, Median, Centroid
- ▶ **Decision Tree:** CART: Prior probabilities, Loss matrix, Split - Information / Gini, Minimum split, Minimum bucket, Complexity parameter, Maximum depth, Cross validation, Plot, Confusion matrix
- ▶ **Neural Network:** Normalization Method - Entropy, Softmax, Censored, Confusion Matrix
- ▶ **Time Series:** Simple Exponential Smoothing, Holt's Method, Holt Winter, ARIMA, ACF Plot, PACF Plot, Time Series Plot
- ▶ **Factor Analysis:** Eigenvalues, Correlations of factor score estimates, Weights, Rotation Matrix, Path diagram, Scree Plot, Parallel Analysis Plot, Correlation Plot
- ▶ **Survival:** Kaplan-Meier, Fleming-Harrington, Nelson - Altschuler
- ▶ **Cox Regression:** Efron, Breslow, Exact, Forest Plot
- ▶ **Charts:** Bar Plot, Scatter Plot, Bar Plot, Box Plot, Pie Chart, Histogram, Heat Map, Pyramid Plot, Dual Axes Plot, Density Plot, Correlogram, Summary Plot, Q-Q Plot
- ▶ **Reports:** Report Tables, Group Summaries
- ▶ **Preprocess:** Derive, Recode, Filler, Lead / Lag, Flag, Filters



SERVER REQUIREMENTS

- ▶ **Hardware:**
 - Processor:** Quad core processor or equivalent
 - Memory:** 16GB RAM or more
 - Hard Disk:** 600 GB or more
- ▶ **Software:**
 - Operating System**
Ubuntu Server 16
 - Database**
PostgreSQL 9.5
 - Web Server**
Apache Tomcat 8.0
 - Java**
JDK 8.0
 - R**
R 3.4.2
- ▶ **Browser:**
Best viewed in Google Chrome
- ▶ **Network:**
TCP/IP protocol
Port 8080 to be opened

Note: Dedicated servers with no shared web applications running