

# **Causal Analytics Toolkit**

## **QuickCAT User Guide**

## System Requirements

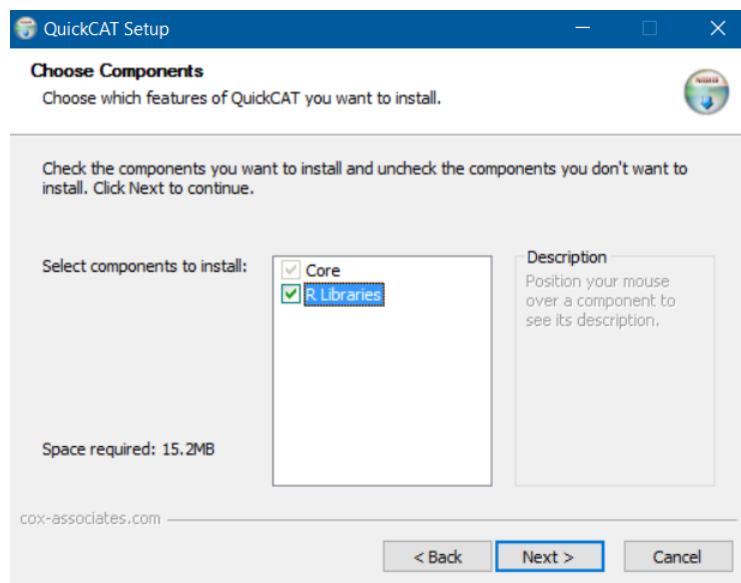
Before installing QuickCAT on Microsoft Windows, the following must be installed:

- Microsoft Excel 2007 or later

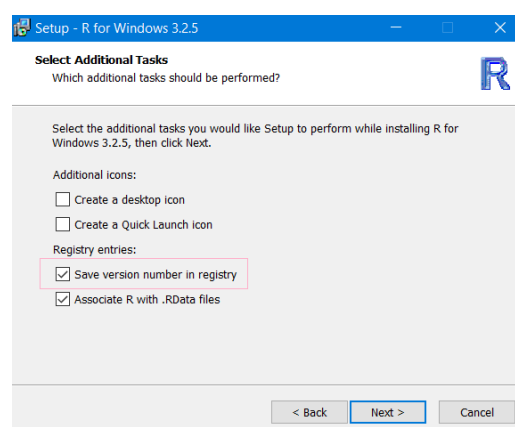
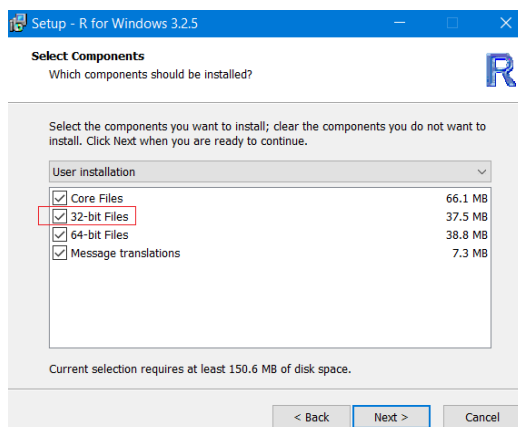
As long as you can run Excel 2007 or later, then you can use QuickCAT.

## Install QuickCAT

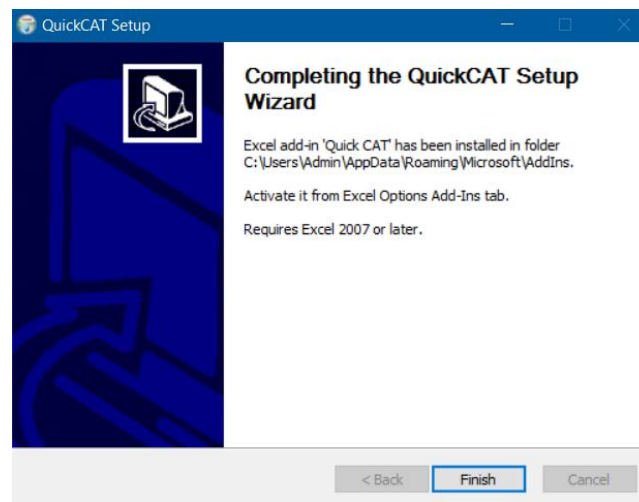
Double-click setup.exe, then follow the on-screen instructions. If you already have R installed, the installer will check whether the installed version can be used by CAT (32-bit R must be installed). If you have not installed R, then the installer will download and install R automatically.



When installing R, options *32-bit Files* and *Save version number in the registry* are checked by default (see below). Do not uncheck these, otherwise CAT will not work.



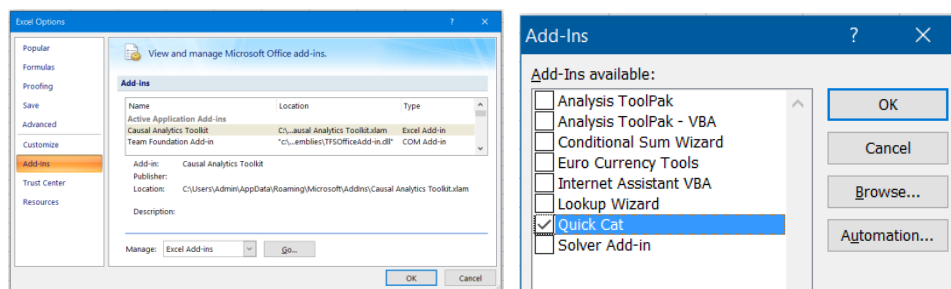
The CAT installer also provides an option to install R libraries (see Figure 1 above). This box must be selected when you install CAT the first time. When installation is complete, an Excel add-in “*Quick CAT.xlam*” will be installed in the default Excel add-in folder. (The target folder can be changed during the installation process if desired). The final screen for the installation process should look as follows:



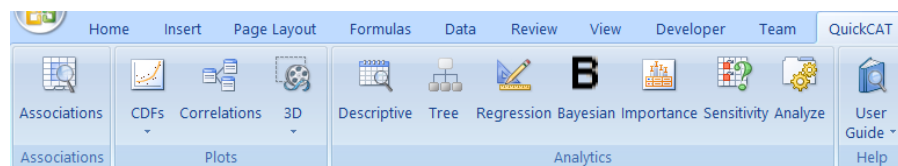
Note: If you manually upgrade R *after* CAT is installed, then you *must* reinstall CAT again.

## Set up QuickCAT Excel add in

Click the top left office icon in Excel (the “Office Button”) then click the “Excel Options” button. Select “Add-ins”, then click the “Go...” button next to Manage Excel Add-ins. Select “Quick CAT”



When the “QuickCAT” appears in the Excel ribbon as shown below, installation is complete.



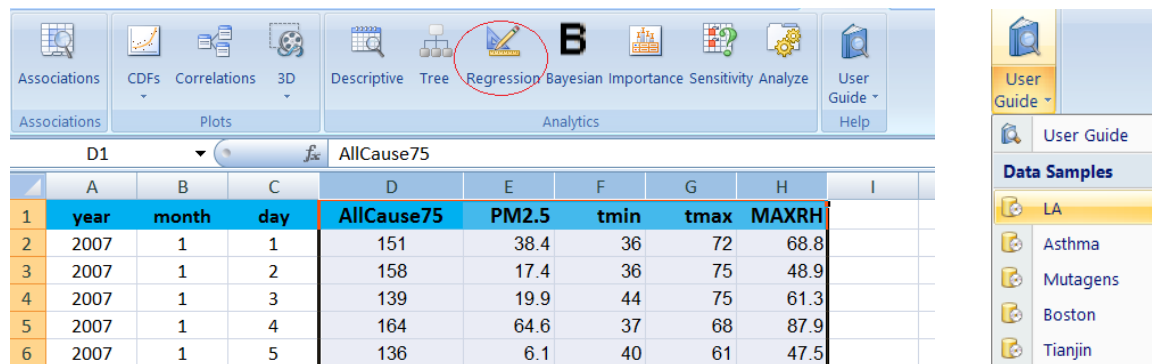
QuickCAT can be uninstalled at anytime using the standard Programs and Features in Windows’ control panel.

## Using QuickCAT

The QuickCAT Excel add-in provides extremely simple, powerful commands and a point-and-click interface for doing advanced analytics from Excel using R packages, even if the user does not know R. For users who have no knowledge of R, a few mouse clicks will display results from advanced R packages without the need to learn R or to know about the packages being used.

### Example: Push-Button Regression Modeling and Bayesian Network Modeling in CAT

To illustrate QuickCAT's push-button analytics, open a new Excel workbook and select Data sample LA from the *User Guide* drop-down menu at the far right of QuickCAT ribbon. A new worksheet named *DataLA* will be created that looks like this:



The screenshot shows the QuickCAT ribbon with various analytics options. The 'Regression' button is highlighted with a red circle. To the right, the 'User Guide' dropdown menu is open, showing 'Data Samples' with 'LA' selected. Below the ribbon, an Excel spreadsheet is shown with the following data:

	A	B	C	D	E	F	G	H	I
1	year	month	day	AllCause75	PM2.5	tmin	tmax	MAXRH	
2	2007	1	1	151	38.4	36	72	68.8	
3	2007	1	2	158	17.4	36	75	48.9	
4	2007	1	3	139	19.9	44	75	61.3	
5	2007	1	4	164	64.6	37	68	87.9	
6	2007	1	5	136	6.1	40	61	47.5	

Highlight columns D-H (by clicking and swiping on the Excel column letters D-H). Click on the Regression button, QuickCAT will then automatically select appropriate families of regression models, fit the models to the data, and display the results. The first column selected, column D (AllCause75) is treated as the dependent variable and the remaining variables highlighted are treated as candidate predictors (i.e., independent variables). CAT generates a new Regression tab with extensive output from the regression modeling, beginning with the following display:

```
CAT_regression (AllCause75,PM2.5,tmin,tmax,MAXRH)
Estimated Coefficients

Call:
glm(formula = fm, family = quasipoisson())

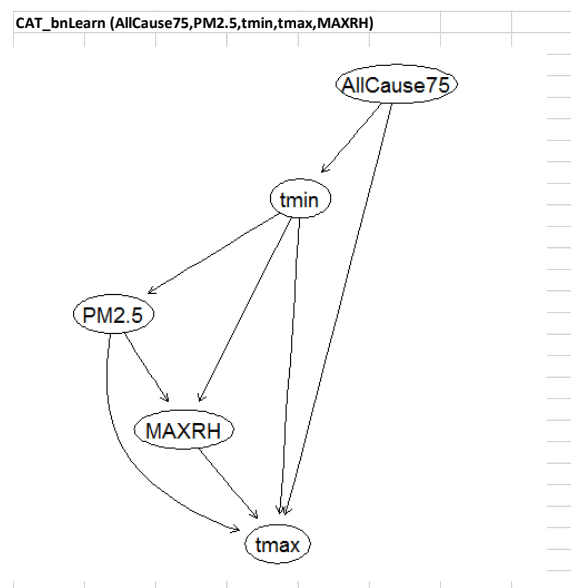
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.7078  -0.9241  -0.0240   0.8433   5.0138

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.3466312  0.0279479  191.307 < 2e-16 ***
PM2.5        0.0005926  0.0002652   2.235  0.0256 *
tmin        -0.0047077  0.0006500  -7.243 7.10e-13 ***
tmax        -0.0018344  0.0004668  -3.930 8.90e-05 ***
MAXRH       -0.0009451  0.0002420  -3.905 9.85e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.611213)
```

This shows the regression coefficients for a quasi-Poisson regression model fit to the data. Additional outputs include 95% confidence intervals, Added-Variable plots showing how the dependent variable is predicted to change as each predictor is varied (i.e., assigned a range of counterfactual values, while holding all other variables at their actual values), importance measures, results from non-parametric (Random Forest) and linear regression modeling, and a plot showing which variables are used in linear regression models of increasing size that explain

increasing proportions of the variance in the dependent variable. These outputs are generated by appropriate R packages and use the same format as these packages; thus, the modeling and interpretation of results can be studied in detail by using the extensive existing documentation on R packages. (To see which specific R packages are used by each CAT function, select *View CAT Functions* under the *Function Builder* drop-down menu in the Methods section of the CAT ribbon.) To generate a Bayesian Network for the same data, simply click on the *Bayesian Network* button (under the Causal Models section of the CAT ribbon). CAT will then generate a new tab called *Bayesian* that shows a Bayesian network (BN) structure for the data. (In such a BN diagram, nodes represent variables and arrows between nodes show that they are not statistically independent of each other, i.e., observing the value of one variable provides information about the value of the other.) CAT does not teach the user how to read and interpret its outputs, leaving this for the R package documentation. But it does enable users to generate a rich set of advanced statistical outputs with minimal effort, and with no knowledge of R or R packages, by selecting columns in an Excel spreadsheet and clicking on analytics buttons.



## Basic Use Cases

### Step 1. Create Data

The input data to QuickCAT must be in a sheet with name prefix *Data*. A few sample data sets are included with the QuickCAT installation. These can be loaded from the *User Guide* drop-down menu. LA is a time series data set for daily fine particulate matter (PM2.5) concentrations, mortality counts among people at least 75 years old (AllCause75), and meteorological variables (daily minimum and maximum temperatures and maximum relative humidity) for the Los Angeles air basin. These data were kindly provided by Dr. Stan Young. Asthma is a larger cross-sectional data set assembled from EPA and CDC ([BRFSS survey](#)) data in the public domain.

Upon clicking a sample data, if there is already a Data sheet for that data set, then it will be just activated. Otherwise, a new worksheet *DataX* will be created, where X is the name of the data set.

To apply QuickCAT to a your data set, open a new workbook in Excel and create an Excel spreadsheet named *DataMydata* that contains the data set, where [Mydata] can be any string. Top row must be

column names. It is suggested to use short column names for easy reports reading. Data should start from the first column, and fill a range without any blank columns in middle.

**Note:** *The current QuickCAT does not allow missing data. If you need the capability to impute missing data, consider CATPro.*

### **Step 2. Using QuickCAT models**

To see results from a QuickCAT model, select some columns from Data worksheet, then click the desired analytics button(s) or icons on the ribbon. The selected columns in the Data sheet are remembered, so you can click on each analytics option in the CAT ribbon after columns in the Data sheet have been selected. Results from running a QuickCAT analytics model are placed into a sheet with a name corresponding to that analysis (e.g., “Correlations” for the results of correlation analyses, “Bayesian” for Bayesian Network, and so forth). The output sheet is cleared and repopulated each time a QuickCAT analytics model is run. To save the results, just rename the output sheet to a sheet with a different name.

## **Conclusion**

This concludes our summary of QuickCAT’s core capabilities. The main purpose of CAT is to give simplified access to the analytics power of the vast array of R packages and commands that are useful for detecting, analyzing, quantifying, and visualizing associations and other relations. No knowledge of R is required.

We encourage users at all levels to use QuickCAT in conjunction with Google. Google can be used to find useful R packages and capabilities. QuickCAT will be updated as R releases and packages are updated and new packages for advanced analytics and causal analysis are added to the CRAN repository. We therefore encourage users to check for QuickCAT updates often [here](#). We also welcome and encourage users to send comments, questions, notifications of bugs or difficulties in using QuickCAT, and suggestions for improvements and additions to [tcxdenver@aol.com](mailto:tcxdenver@aol.com).

We hope to make QuickCAT and its add-on modules as useful and easy-to-use as possible. Your feedback will help to achieve this goal.