Exploratory Only: A Tool for Large-Scale Exploratory Analyses

Jin Kim¹

¹ Yale University, School of Management

Author Note

Jin Kim https://orcid.org/0000-0002-5013-3958

Data, videos, and additional online materials are openly available at the project's Open Science Framework page (https://osf.io/xcq2n/). I have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Jin Kim, School of Management, Yale University, 165 Whitney Avenue, New Haven, CT 06511, United States. Email: jin.m.kim@yale.edu

Abstract

This article presents Exploratory Only: an intuitive tool for conducting large-scale exploratory analyses easily and quickly. Available in three forms (as a web application, standalone program, and R Package) and launched as a point-and-click interface, Exploratory Only allows researchers to conduct all possible correlation, moderation, and mediation analyses among selected variables in their data set with minimal effort and time. Compared to a popular alternative, SPSS, Exploratory Only is shown to be orders of magnitude easier and faster at conducting exploratory analyses. The article demonstrates how to use Exploratory Only and discusses the caveat to using it. As long as researchers use Exploratory Only as intended—to discover novel hypotheses to investigate in follow-up studies, rather than to confirm nonexistent a priori hypotheses (i.e., *p*-hacking)—Exploratory Only can promote progress in behavioral science by encouraging more exploratory analyses and therefore more discoveries.

Keywords: exploratory analysis, confirmatory analysis, automation, point-and-click interface, *p*-hacking

Exploratory Only: A Tool for Large-Scale Exploratory Analyses Introduction

Exploratory analyses are not inherently bad. Despite their negative association with *p*-hacking (Nelson et al., 2018; Simmons et al., 2011), exploratory analyses serve an important function of generating hypotheses for future research. Insofar as they are a form of *exploration*, exploratory analyses may even be considered an "essential component of science" that promotes "new discoveries and scientific progress," as Wagenmakers et al. concede (2012, p. 635). If this claim is too much to swallow, then most researchers would at least agree with the claim that exploratory analyses *can* lead to novel hypotheses and thereby advance science. It is with this latter claim that this article presents *Exploratory Only*: a tool for conducting large-scale exploratory analyses. In the rest of this section, I briefly explain what Exploratory Only is, why it was developed, and define the term *exploratory analyses* as used in this article.

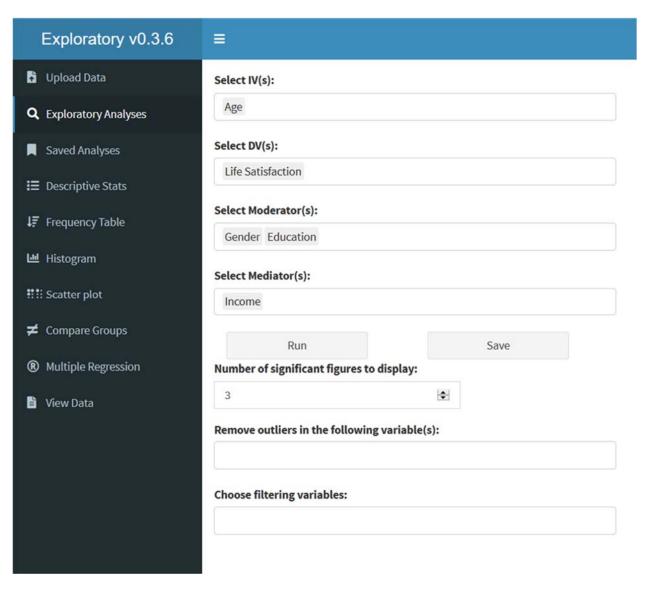
What Is Exploratory Only?

Put simply, Exploratory Only is a tool that automates large-scale exploratory analyses. Exploratory Only provides an intuitive, point-and-click graphical user interface (GUI; see Figure 1) powered by R's Shiny App framework (Chang et al., 2021; Chang & Borges Ribeiro, 2018) and other R packages (Behrendt, 2014; Dowle & Srinivasan, 2021; Edwards, 2020; Tingley et al., 2014; Wickham, 2016; Wilke, 2021; Xie et al., 2021). With Exploratory Only, researchers can intuitively (1) upload their data set, (2) select variables of interest, and (3) conduct with one click a set of all possible correlation, moderation, and mediation analyses. Available in three forms—as a web application, standalone program, and R package—Exploratory Only allows researchers to conduct exploratory analyses more exhaustively, more easily, and more quickly than alternative tools. Later in this article, I demonstrate the convenience and speed of

Exploratory Only by comparing it with a benchmark, Statistical Package for the Social Sciences (SPSS; IBM Corp., 2020), a statistical analysis tool widely popular among researchers (Muenchen, 2019).

Figure 1

Intuitive User Interface of Exploratory Only



Why Was Exploratory Only Developed?

I developed Exploratory Only for several reasons. First, as alluded to earlier, I suspected that exploratory analyses may have fallen out of fashion with "Psychology's Renaissance" (Nelson et al., 2018) and felt that a tool like Exploratory Only could help bring exploratory analyses back into vogue. Following the notable events that started Psychology's Renaissance (Nelson et al., 2018), such as the publication of Daryl Bem's article on extrasensory perception (Bem, 2011), the revelation of Diederik Stapel's scientific fraud (Levelt et al., 2012), and the publication of Simmons, Nelson, and Simonsohn's seminal article on false-positive psychology (Simmons et al., 2011), much of researchers' attention shifted toward correct ways of conducting confirmatory analyses and perhaps away from exploratory analyses. Certainly, we as researchers should conduct confirmatory analyses properly (i.e., not p-hack), but also, we should not neglect to conduct exploratory analyses, because they can generate new ideas and hypotheses to investigate. The influential statistician John Tukey (Kafadar, 2001) noted that "ideas come from previous exploration more often than from lightning strokes," and therefore "extensive exploration (of past data) can—and should—play a role in ... formulating the question" (Tukey, 1980). Exploratory Only responds to Tukey's call by providing a user-friendly and convenient tool (as we shall see) and assisting researchers with more extensive exploration of their data. If, as I suspect, there is currently an imbalance heavily favoring confirmatory analyses over exploratory analyses in the field of behavioral science, mitigating the imbalance may benefit the field, because "neither exploratory nor confirmatory is sufficient alone," and we "need them both" (Tukey, 1980). Thus, Exploratory Only was developed partly to address this imbalance and to remind researchers of the need for exploratory analyses.

6

Other reasons for developing Exploratory Only were more practical. The current practice of exploratory analyses may be tedious, time-consuming, and unsystematic, and Exploratory Only was developed to improve this. Often, researchers focus mostly on exploratory analyses that jump out at them or make sense to them. For example, if a previously overlooked Measure X in a study's data set is similar to the dependent measure of the study, a researcher might conduct exploratory analyses with Measure X as a dependent measure. However, the researcher may fail to consider conducting exploratory analyses with another overlooked Measure Y because it seems unrelated to other variables of interest—despite the possibility that including Measure Y in the exploratory analyses could produce thought-provoking results. Alternatively, if the researcher did attempt to conduct exploratory analyses more systematically and exhaustively, they may have found the process to be tedious and time-consuming, because they would have had to test so many relationships among variables of interest. As in this example, researchers may conduct exploratory analyses only superficially, skip potentially interesting analyses unintentionally, or even forgo exploratory analyses altogether because of the constraints in time and effort. Unfortunately, then, results from exploratory analyses that may lead to fruitful investigation in follow-up studies would end up tucked away in the "file drawers" (Rosenthal, 1979) of researchers. Exploratory Only can improve the current state of exploratory analyses. By making exploratory analyses not only more exhaustive, but also faster and easier, Exploratory Only can uncover interesting relationships that may have been overlooked otherwise and generate novel hypotheses for future research.

Defining Exploratory Analyses as Used in This Article

Before examining Exploratory Only more closely, it would be helpful to define the term *exploratory analyses* more precisely, as used in this article. As in previous research, *exploratory*

analyses in this article refer to analyses that generate hypotheses, as opposed to confirmatory analyses that confirm hypotheses (Tukey, 1980; Wagenmakers et al., 2012). More specifically, however, *exploratory analyses* in this article refer to a set of all possible (1) correlation, (2) moderation, and (3) mediation analyses that can be conducted among variables of researcher's interest in the data set (unless noted otherwise). For example, if a researcher thinks that Variable A in their data set can act as an independent variable, B a moderating variable, C a mediating variable, and Y and Z dependent variables, then *exploratory analyses* will refer to a set of 2 + 2 + 2 = 6 possible analyses as follows: 2 possible correlation analyses (1 independent variable × 2 dependent variables); 2 possible moderation analyses (1 independent variable × 1 moderator variable × 2 dependent variables); and 2 possible mediation analyses (1 independent variable × 1 mediator variable × 2 dependent variables). In this article, then, *exploratory analyses* refer to an *exhaustive* set of all possible correlation, moderation, and mediation analyses, whereas exploratory analyses in other research articles usually refer to a *selective* set of analyses that researchers have in mind (often, just a small number of them).

Overview of the Paper

The rest of this article is organized as follows. First, I demonstrate the convenience and speed of Exploratory Only by comparing it with the benchmark SPSS. I then show how researchers can use Exploratory Only available in three forms. Next, I introduce basic analysis functions that come with Exploratory Only. Finally, I conclude the article with the caveat that any results obtained from Exploratory Only should be treated merely as exploratory and should not be reported as confirmatory.

Comparison of Convenience

The biggest advantage of Exploratory Only over SPSS is the convenience and ease with which large-scale exploratory analyses can be conducted. The convenience and ease come from two central features of Exploratory Only: (1) systematic listing of all possible analyses to conduct, and (2) automatic conducting of all such analyses with one click. I discuss each of these two features below.

Systematic Listing of All Possible Analyses

Exploratory Only brings convenience first by performing the tedious job of systematically listing all possible analyses to conduct. One of the first steps in exploratory analyses is to determine which variables can be candidates for independent, dependent, moderating, and mediating variables. This is a step that researchers must perform themselves, and neither Exploratory Only nor SPSS can perform it for them (though researchers using Exploratory Only can mindlessly enter all variables of interest to be each type of variable). Next, researchers need to consider all possible relationships among the variables of each type and create a list of analyses to conduct. Without this list, researchers in the midst of exploratory analyses will find it difficult to keep track of which analyses have already been conducted and which still need to be conducted. Researchers relying on SPSS will need to manually create this list of analyses, which can be a tedious task. For example, a researcher conducting exploratory analyses for 3 independent variables, 4 dependent variables, 5 moderator variables, and 2 mediator variables will need to create a list of $3 \times 4 = 12$ correlation analyses, $3 \times 4 \times 5 = 60$ moderation analyses, and $3 \times 4 \times 2 = 24$ mediation analyses, for a total of 12 + 60 + 24 = 96analyses—not a pleasant task without programming. In contrast, researchers using Exploratory Only does not have to worry about this task. Exploratory Only creates the list of analyses to

conduct by generating all possible combinations among variables of different types and then eliminating combinations that would not make sense (e.g., a mediation relationship having the same variable as both the independent and dependent variables will be eliminated). Thus, Exploratory Only takes over the tedious task of listing all analyses to conduct.

Automatic Conducting of All Analyses With One Click

Exploratory Only then makes exploratory analyses easier by automatically conducting all analyses with one click. On SPSS, researchers need to take several distinct actions to conduct one analysis. For example, to conduct a simple correlation analysis, a researcher needs to take 3 actions to open the analysis dialog (i.e., navigating to "Analyze" \rightarrow "Correlate" \rightarrow "Bivariate"), 2-4 actions to select two variables of interest (i.e., choosing two variables and moving them to the "Variables" section), and 1 additional action to conduct the analysis (i.e., clicking "OK"). Moreover, these 6-8 actions must be repeated for each of dozens or hundreds of analyses, and this would amount to hundreds or thousands of repetitive actions. In contrast, on Exploratory Only, researchers can select variables of interest just once for all analyses and click the "Run" button once to automatically conduct all dozens or hundreds of analyses. In the following section, I show how Exploratory Only makes exploratory analyses not only easier and more convenient but also faster.

Comparison of Speed

Exploratory Only was markedly faster than SPSS in performing large-scale exploratory analyses. To compare the speed of Exploratory Only and SPSS, I first constructed "small," "medium," and "large" data sets containing 100, 1,000, and 10,000 observations, respectively (Kim, 2021). Each data set consisted of six randomly generated variables having 2, 3, 5, 7, 11, and 101 levels, respectively. That is, in each data set, the first variable, vI, was created by taking

100, 1,000, or 10,000 random samples from the set $\{1, 2\}$ with replacement; the second variable, v2, was created by taking 100, 1,000, or 10,000 random samples from the set $\{1, 2, 3\}$ with replacement; and, likewise, the remaining four variables were created by taking 100, 1,000, or 10,000 random samples with replacement from the set $\{1, 2, ..., k\}$, where k was the number of levels in the variable (i.e., 2, 3, 5, 7, 11, or 101). These three data sets and the R code with the seed value for constructing them are available on this article's page at the Open Science Framework (OSF), https://osf.io/xcq2n/ (hereafter, simply "the OSF").

In each set of exploratory analyses—correlation, moderation, and mediation analyses— Exploratory Only was markedly faster than SPSS. Table 1 summarizes these results.

Table 1

Comparison of Speed, SPSS With PROCESS vs. Exploratory Only

Data Set	Estimated Execution Time (seconds)		Speed
	SPSS with PROCESS	Exploratory Only	Improvement Factor
Correlation Analysis (Number of Analyses = 15)			
N = 100	67.5	7.1	9.5
N = 1000	67.5	6.7	10.1
N = 10000	67.5	7.8	8.7
Moderation Analysis (Number of Analyses = 60)			
N = 100	464	9.6	49
N = 1000	486	10.5	47
N = 10000	490	10.6	46
Mediation Analysis (Number of Analyses = 120)			
N = 100	876	235	3.7
N = 1000	1052	350	3.0
N = 10000	1284	875	1.5
All Analyses Combined (Number of Analyses = 195)			
N = 100	1408	236	6.0
N = 1000	1606	332	4.8
N = 10000	1842	883	2.1

Correlation Analyses

Exploratory Only was about 8 to 10 times faster than SPSS in conducting correlation analyses for our three data sets. Among the 6 variables in each data set, a total of 6 choose 2 = 15 pairwise correlations could be calculated. Thus, I conducted a total of 15 correlation analyses, separately on SPSS and Exploratory Only. On SPSS, I calculated correlations by using the same function ("Analyze" \rightarrow "Correlate" \rightarrow "Bivariate" \rightarrow "Pearson") and manually deselecting and

selecting a pair of variables 15 times. As shown on videos available at the OSF, each analysis took about 4-5 seconds to complete—even with keyboard shortcuts—regardless of the size of our data sets. Much of this time was spent on clicking buttons, rather than on actual computations of correlations, which were completed instantly. I thus estimated that conducting the 15 correlations on SPSS would take about $4.5 \times 15 = 67.5$ seconds, for all three data sets (Table 1, Rows 1-3, Column 1). The process was much simpler and faster on Exploratory Only: I simply entered all 6 variables as both independent and dependent variables and clicked the "Run" button once to calculate all 15 correlations. As shown on videos available at the OSF, the whole process took only about 6-8 seconds, regardless of the size of our data sets (Table 1, Rows 1-3, Column 2). Here, too, much of the time was spent on clicking buttons, rather than on computation of individual correlations, which was completed in a second or two.

Moderation Analyses

Exploratory Only was faster than SPSS by at least an order of magnitude (46 to 49 times) in conducting moderation analyses for our three data sets. Among the 6 variables in each data set, I could construct a total of $6 \times 5 \times 4 = 120$ moderation models: the independent variable of a moderation model could be any of the 6 variables; the dependent variable could be any of the remaining 5 variables; and the moderator variable could be any of the still remaining 4 variables. However, half of these models were redundant, because a moderation model is statistically identical whether or not the independent and moderator variables are switched. For example, a moderation model of age predicting income with sex as a moderator, $Income = b_0 + b_1$ Age $+ b_2$ Sex $+ b_3$ (Age \times Sex) + e, is statistically identical to a moderation model of sex predicting income with age as the moderator, $Income = b_0 + b_1$ Sex $+ b_2$ Age $+ b_3$ (Sex \times Age) + e. I thus needed to test only the nonredundant half of the moderation models, or 60 moderation models.

13

On SPSS, I used PROCESS Model 1 (Hayes, 2017) to estimate the time required to conduct 60 moderation analyses. Instead of using PROCESS Model 1, I could have manually computed the interaction term in each moderation model (by navigating to "Transform" → "Compute Variable") and used SPSS's built-in linear regression function ("Analyze" \rightarrow "Regression" \rightarrow "Linear"). However, this method would have taken more steps and time, and therefore the current PROCESS Model 1 method is a more conservative benchmark for Exploratory Only. As shown on videos available at the OSF, each analysis with PROCESS Model 1 took about 7-9 seconds to complete (even with keyboard shortcuts), regardless of the size of our data sets. When I measured the time it took to test one specific model (v1 predicting v2, moderated by v4), results for the three data sets were 7.74, 8.10, and 8.17 seconds respectively. I multiplied these times by 60 to estimate that conducting the 60 moderation analyses for small, medium, and large data sets would take 468, 486, and 490 seconds, respectively (Table 1, Rows 4-6, Column 1). The difference in the estimated times may be due simply to random differences in clicking speed, rather than due to systematic differences in computation load stemming from sizes of our data sets.

On Exploratory Only, conducting 60 moderation analyses was much simpler and faster. I simply entered all 6 variables as independent, dependent, and moderator variables and clicked the "Run" button once to calculate the interaction term's *p*-value for all 60 moderation models. As shown on videos available at the OSF, the whole process took only 9.6, 10.5, and 10.6 seconds for small, medium, and large data sets, respectively (Table 1, Rows 4-6, Column 2). Size of our data sets did not seem to impact the execution time much.

Mediation Analyses

Exploratory Only was about 1.5 to 3.7 times faster than SPSS in conducting mediation analyses for our three data sets. Among the 6 variables in each data set, I could construct a total of $6 \times 5 \times 4 = 120$ mediation models: the independent variable of a mediation model could be any of the 6 variables; the dependent variable could be any of the remaining 5 variables; and the mediator variable could be any of the still remaining 4 variables. Unlike in moderation analyses, all 120 mediation models were statistically unique.

On SPSS, I used PROCESS Model 4 (Hayes, 2017) to estimate the time required to conduct 120 mediation analyses. As shown on videos available at the OSF, each analysis took about 7-11 seconds to complete. Unlike in correlation or moderation analyses, size of our data sets had noticeable impacts on the execution time. When I measured the time it took to test one specific model (*v1* having an indirect effect on *v2* through *v4*), results for the three data sets were 7.30, 8.77, and 10.70 seconds respectively. I multiplied these times by 120 to estimate that conducting the 120 mediation analyses for small, medium, and large data sets would take 876, 1052, and 1284 seconds, respectively (Table 1, Rows 7-9, Column 1).

On Exploratory Only, conducting 120 mediation analyses was much simpler and quite a bit faster, though the improvement in speed was not as impressive as that with correlation or moderation analyses. I conducted these mediation analyses by simply entering all 6 variables as independent, dependent, and mediator variables and clicking the "Run" button once to calculate the *p*-value of the indirect effect for all 120 mediation models. Exploratory Only calculates these *p*-values of indirect effects using the "mediate" function from Tingley and colleagues' R package "mediation" (Tingley et al., 2014). As shown on videos available at the OSF, mediation analyses took significantly longer for bigger data sets: 235, 350, and 875 seconds for small, medium, and

large data sets, respectively (Table 1, Rows 7-9, Column 2). Although these execution times were much longer than those for correlation or moderation analyses, they were still a big improvement from the estimated execution times for SPSS.

All Analyses Combined

Exploratory Only was about 2 to 6 times faster than SPSS in conducting all three sets of exploratory analyses. To estimate the total execution time for all 195 analyses for SPSS, I simply added up the estimated execution times for correlation, moderation, and mediation analyses (e.g., for the small data set, 67.5 + 464 + 876 = 1408); see Table 1, Rows 10-12, Column 1. As for Exploratory Only, I measured the total execution time for all 195 analyses in each data set, as shown on videos available at the OSF. Again, as with mediation analyses, the total execution time on Exploratory Only increased with size of our data sets: 236, 332, and 883 seconds for small, medium, and large data sets (Table 1, Rows 10-12, Column 2). More importantly, however, conducting exploratory analyses was much faster with Exploratory Only than with SPSS for each data set.

Speed Comparison in Reality

That Exploratory Only was faster than SPSS by a factor of 2 to 6, however, is a gross understatement. In reality, it would be unreasonable to think that researchers can conduct all 195 analyses on SPSS within 1400 to 1800 seconds (about 23 to 30 minutes), as estimated here. Most likely, researchers will be unable to keep the constant pace in a series of repetitive analyses; will get tired from too much clicking and need to take breaks; and will make mistakes by misclicking or repeating some of the same analyses. Staying alert, accurate, systematic, and motivated through dozens or hundreds of repetitive analyses is not an easy task. Moreover, combing through pages of SPSS outputs, copying and pasting *p*-values of interest, and tabulating variables

and results from all analyses will also take a long time. If all these factors are taken into account, Exploratory Only may be orders of magnitude faster than SPSS in conducting large-scale exploratory analyses.

Recommended Usage

Exploratory Only is available in three different forms: as a web application, a standalone program, or an R package. Across the three forms, researchers can use Exploratory Only the same way. I first demonstrate how to use the web application and discuss how the standalone program and the R package can be downloaded and set up.

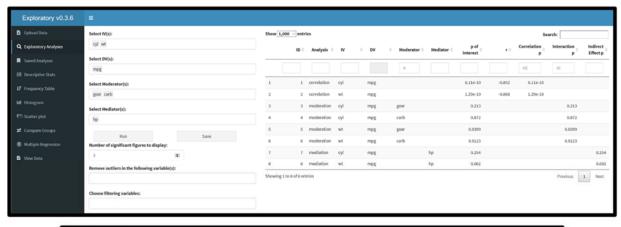
Web Application (and How to Use Exploratory Only)

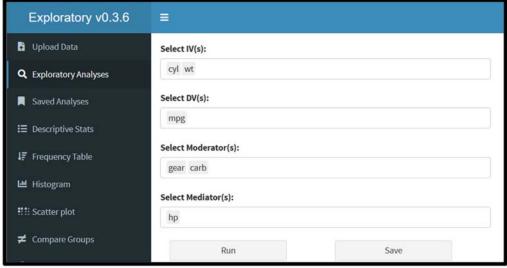
As of April 2021, Exploratory Only is available as a web application through a link on this article's website, https://exploratoryonly.com. To try using Exploratory Only, researchers are encouraged to visit the website and click on the link for "Web Application," which will launch Exploratory Only in a new tab on the current browser. As shown on the top panel of Figure 2, the interface of Exploratory Only is intuitive enough that explicit instructions on how to use it would not be necessary. Researchers can try using Exploratory Only with the sample data that had been preloaded, the "mtcars" data set that comes with base R (R Core Team, 2021). All researchers need to do is select variables as independent, dependent, moderating, and mediating variables, and click on "Run" button to conduct all possible analyses involving all combinations of the variables selected. For example, as shown in the middle panel of Figure 2, researchers may select "cyl" (number of cylinders) and "wt" (weight) as independent variables, "mpg" (miles per gallon) as a dependent variable, "gear" (number of forward gears) and "carb" (number of carburetors) as moderating variables, and "hp" (gross horsepower) as a mediating variable. Clicking on "Run" button then conduct a total of 8 analyses: 2 correlation, 4 moderation, and 4

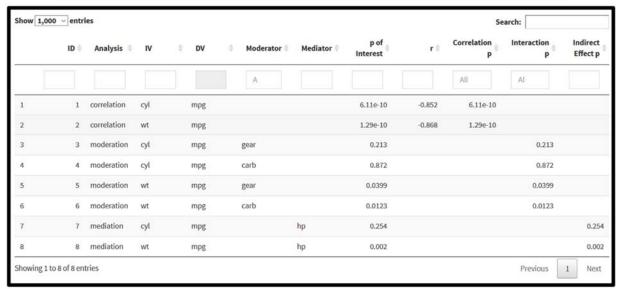
17

mediation analyses. As shown in the bottom panel of Figure 2, output of the exploratory analyses will be a table listing the eight analyses (ID: 1-8), types of analysis, variables by types, and *p*-values of interest (i.e., *p*-value of the correlation coefficient in correlation analyses, *p*-value of the interaction term in moderation analyses, and *p*-value of the indirect effect in mediation analyses).

Figure 2
Using Exploratory Only With Sample Data







Once researchers have an idea of how Exploratory Only works with sample data, they are encouraged then to upload their own data and conduct exploratory analyses of their own. To do so, researchers can click on "Upload Data" on the sidebar menu, browse for their data file (a .csv file), and upload it. The maximum size of the data file that can be uploaded is 50 MB, but researchers can work with a larger data file by using the standalone program or the R package discussed next.

Standalone Program

Exploratory Only is also available as a standalone program for Microsoft Windows. Researchers can download it either from the article's website, https://exploratoryonly.com, or from its Open Science Framework page (https://exploratoryonly.com, or from its Open Science Framework page (https://exploratoryonly.com, or from its Open Science Framework page (https://osf.io/xcq2n/). After downloading the file, researchers should first extract the downloaded file ("Exploratory Only v0_3_8.zip"), which takes about 2 minutes and 10 seconds. Once the extraction is completed, researchers can navigate to the folder with extracted files and execute the Visual Basic Script file, "Exploratory Only (Execute This File).vbs." This will launch Exploratory Only on the default web browser of a local machine, and researchers can use it the same way as with the web application (see the previous Web Application section).

R Package

Running the Exploratory Only in R is also straightforward. The R Package of Exploratory Only is called "exploratory," and it is available for download on the Comprehensive R Archive Network (CRAN; Kim, 2021b) and the author's GitHub repository (https://github.com/jinkim3/exploratory), which is also linked on the article's website, https://exploratoryonly.com. Researchers can download the most recent version of the package with two lines of code in R (Section 1 of the Appendix). Once the package is installed,

researchers can simply type one line of code to launch Exploratory Only (Section 2 of the Appendix). Researchers are encouraged to periodically update the package when they are notified of new version releases (Section 3 of the Appendix). In short, running Exploratory Only in R is extremely simple: just two lines of code to download and install the package, and one line of code to launch the tool on a web browser of the local machine (Appendix). Once Exploratory Only is launched from R, researchers can use it the same way as with the web application (see the previous Web Application section).

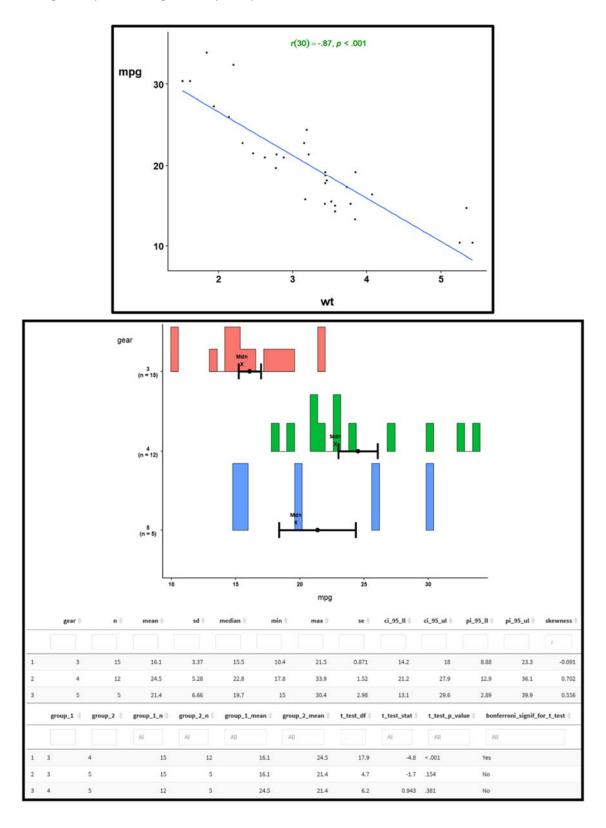
Other Analysis Functions

Exploratory Only also provides functions to conduct basic analyses with a point-andclick method. These functions can be accessed by clicking the sidebar menu of Exploratory Only (left side of Figure 1). Upon discovering potentially interesting relationships from exploratory analyses, researchers can use these functions to quickly conduct simple follow-up—but not confirmatory—analyses. For example, if a researcher comes across an unexpected correlation between Variables A and B, they can conduct a follow-up analysis by (1) clicking on "Scatter Plot" tab on the left; (2) selecting Variable A as the independent variable and Variable B as the dependent variable (or vice-versa); and (3) clicking on the "Run" button. This will produce a scatter plot showing the relationship between the two variables both graphically (observed values of the two variables and a regression line) and statistically (correlation estimate and the associated p-value), as shown in the top panel of Figure 3. Likewise, researchers can follow up on particular results from exploratory analyses with pairwise t-tests and Mann-Whitney tests, as well as estimation of effect sizes (bottom panel of Figure 3) by navigating to the "Compare Groups" tab. In addition, researchers can test multiple regression models with up to 3-way interactions on the "Multiple Regression" tab or examine single variables more closely on

"Descriptive Stats," "Frequency Table," or "Histogram" tabs. These functions are incorporated from the R packages, *ezr* and *kim*, by the same author, and Exploratory Only will be updated periodically with newer and more functions from both packages (Kim, 2020, 2021c).

Figure 3

Follow-Up Analyses on Exploratory Only



Caveat

The biggest caveat to using Exploratory Only is that results from it should never be reported as confirmatory. If Exploratory Only reveals any relationship that might be worth investigating (e.g., a significant correlation between two variables whose link had not been hypothesized or obvious ex ante), researchers must conduct follow-up studies to put it through a proper empirical test before reporting it. Exploratory Only will find a statistically significant relationship in about 5% of the analyses at $\alpha = .05$ even among randomly generated variables that should have no relationship whatsoever with one another. What is more, this 5% is likely to be a lower bound, because factors such as common method bias could inflate the false discovery rate (Bagozzi & Yi, 1993; Campbell & Fiske, 1959; Cote & Buckley, 1987; Podsakoff et al., 2003; but see Spector, 2006 for opposing views). Moreover, because Exploratory Only is intended to conduct analyses in the order of 100s or 1000s, researchers may discover dozens or even hundreds of significant relationships. That is, even with methods controlling for false discovery rates (Benjamini & Hochberg, 1995), some relationships may reach statistical significance due to the sheer number of hypotheses being tested.

Before reporting any such relationship as confirmatory, however, researchers must take proper steps of empirical research: formally hypothesize the relationship, (ideally) preregister the hypothesis, and empirically test the hypothesis in a rigorous manner. Skipping this essential empirical test and reporting the relationship as confirmatory or as supporting an a priori hypothesis (which in fact never existed) would be grossly unscrupulous and amount to scientific fraud. Evidence for the relationship could very well be nothing more than a statistical artifact obtainable only through an inordinate number of analyses. Claiming to have hypothesized and "found" the relationship, or claiming to confirm the hypothesis (which never existed a priori)

will mislead the audience and hinder the progress of science, and therefore should be avoided at all costs.

Conclusion

This article presents Exploratory Only, a user-friendly tool that makes exploratory analyses more exhaustive, easier, and faster than popular alternatives like SPSS. With Exploratory Only, researchers can explore their data more thoroughly with minimal time and effort and possibly discover novel hypotheses or potentially interesting relationships among variables that otherwise would have remained buried in their data. As long as researchers remember that results from Exploratory Only are never confirmatory and must be properly tested in follow-up studies, Exploratory Only can promote progress in behavioral science by encouraging more exploratory analyses and therefore more discoveries.

Appendix

R Code to Install the R Package "exploratory" and Launch Exploratory Only

1. Install the most recent version of the package from GitHub. install.packages("remotes") remotes::install_github("jinkim3/exploratory")

2. Launch Exploratory Only on the local machine. exploratory::exploratory()

3. Update the package periodically. update_exploratory()

Open Practices Statement

The data, videos, and additional online materials are openly available at the project's Open Science Framework page (https://osf.io/xcq2n/). No experiments were conducted for this article, and therefore no experiment was preregistered.

References

- Bagozzi, R. P., & Yi, Y. (1993). Multitrait–multimethod matrices in consumer research: Critique and new developments. *Journal of Consumer Psychology*, 2(2), 143–170. https://doi.org/10.1016/S1057-7408(08)80022-8
- Behrendt, S. (2014). *lm.beta: Add Standardized Regression Coefficients to lm-Objects*. https://cran.r-project.org/package=lm.beta
- Bem, D. J. (2011). Feeling the Future: Experimental Evidence for Anomalous Retroactive

 Influences on Cognition and Affect. *Journal of Personality and Social Psychology*, 100(3),
 407–425. https://doi.org/10.1037/a0021524
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B* (Methodological), 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016
- Chang, W., & Borges Ribeiro, B. (2018). *shinydashboard: Create Dashboards with "Shiny."* https://cran.r-project.org/package=shinydashboard
- Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2021). *shiny: Web Application Framework for R*. https://cran.r-project.org/package=shiny
- Cote, J. A., & Buckley, M. R. (1987). Estimating Trait, Method, and Error Variance:

 Generalizing across 70 Construct Validation Studies. *Journal of Marketing Research*, 24(3), 315. https://doi.org/10.2307/3151642

- Dowle, M., & Srinivasan, A. (2021). *data.table: Extension of `data.frame*`. https://cran.r-project.org/package=data.table
- Edwards, S. M. (2020). *lemon: Freshing Up your "ggplot2" Plots*. https://cran.r-project.org/package=lemon
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.
- IBM Corp. (2020). IBM SPSS Statistics for Windows (27.0).
- Kafadar, K. (2001). *John W. Tukey (1915–2000)*. Amstat News. https://magazine.amstat.org/blog/2001/06/16/tukey-sih/
- Kim, J. (2020). ezr: Easy Use of R via Shiny App for Basic Analyses of Experimental Data (R Package version 0.1.5). The Comprehensive R Archive Network, CRAN.
 https://doi.org/10.5281/zenodo.5140692
- Kim, J. (2021a). Data Sets for Speed Comparison Between Exploratory Only and SPSS (Version 1). https://osf.io/xcq2n/
- Kim, J. (2021b). exploratory: A Tool for Large-Scale Exploratory Analyses (R Package version 0.3.13). The Comprehensive R Archive Network, CRAN. https://doi.org/10.5281/zenodo.5140699
- Kim, J. (2021c). kim: Behavioral Scientists' Analysis Toolkit (R Package version 0.3.72). The Comprehensive R Archive Network, CRAN. https://doi.org/10.5281/zenodo.4445388
- Levelt, W. J. M., Drenth, P. J. D., & Noort, E. (2012). Flawed science: The fraudulent research practices of social psychologist Diederik Stapel.
- Muenchen, R. A. (2019). *The Popularity of Data Science Software*. http://r4stats.com/articles/popularity/

- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology 's Renaissance.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant.

 Psychological Science, 22(11), 1359–1366. https://doi.org/10.1177/0956797611417632
- Spector, P. E. (2006). Truth or urban legend? Organizational Research Methods, 9(2), 221–232.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5), 1–38. https://doi.org/10.18637/jss.v059.i05
- Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician*, 34(1), 23–25.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012).

 An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, 7(6), 632–638. https://doi.org/10.1177/1745691612463078
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wilke, C. O. (2021). *ggridges: Ridgeline Plots in "ggplot2."* https://cran.r-project.org/package=ggridges

Xie, Y., Cheng, J., & Tan, X. (2021). *DT: A Wrapper of the JavaScript Library "DataTables."* https://cran.r-project.org/package=DT