eRg
www.ergreports.com

Re: Partner Data Reporting Contractor (TRAILS)

Dear contractor selection team,

I am writing in response to your recent Partner Data Reporting Contractor posting. I was, for the majority of my career, a full-time professor of Industrial and Organizational Psychology. Currently I am the primary care provider for an ailing relative, but I also maintain a small LLC (eRg) that is intended to serve project-based contracts such as yours. The company (eRg) and the person (John Kulas) are essentially the same entity – I am not currently pursuing any contracts other than yours.

My primary areas of specialization have always been centered on **work processes that support the integrity and fidelity of data**. In 2008 I published a book that frames the principles of proper data management practices, although the most recent (2nd) edition of that text reads more like a standard undergraduate statistics guide (because the publisher requested that shift in focus). I do also have over 20 years of experience "communicating complex data topics with diverse audiences" including apprehensive students, highly competent researchers, and (occasionally data-averse) Human Resource professionals.

Although I am no longer a full-time professor, I am still interested in providing guidance regarding good data management practices – particularly within the framework of `R`. I have several resources on my website and also maintain a YouTube channel with a weekly "Open Office Hours" LiveStream that is intended to help others develop reproducible research practices via inclusive specification of prose and analyses with `R`, `Rmarkdown`, and `Quarto`.

Regarding Qualtrics data, I typically rely on base `R` functions rather than slicker package functions (such as, for example, `tidyverse` options) because function conflicts are more likely to covertly arise when packages are utilized. My code is therefore not exceedingly *efficient* in terms of length, but it is structured in a manner that minimizes the likelihood of hidden/unknown errors in data processing. Comments are used copiously and always include a date.

I typically access Qualtrics data by "chopping" the top 2 rows (see line 3,

below), retaining one of those rows for variable name specification (line 5), and then adding those variable names to the previously "chopped" dataframe (line 7):

```
## Cleaning Qualtrics construct validation data - 10/14/21

data.att <- read.csv("October+12,+2021_08.02.csv")[-c(1:2),]

varnames <- read.csv("October+12,+2021_08.02.csv")[1,c(1:92)]

names(data.att) <- varnames
```

Regarding data cleaning, my personal surveys typically have quality-check items and I employ case-wise deletion. In addition to explicit quality-checks, I also compute and evaluate Qualtrics survey durations and truncate respondents with unreasonably small response durations:

If the survey is "long enough", I will also typically assess consecutive non-differentiating responses as well as response variability – the treatment of extreme cases here is context-dependent, but always involves solicitation of a frequency distribution:

```
library(careless)
newdata.att$careless_long <- longstring(newdata.att[20:83])
newdata.att$irv <- irv(newdata.att[48:51])  ## same response scale
```

References

Name	Relation	Contact.Info
Alicia Stachowski (PhD)	Collaborator (Research)	stachowskia@uwstout.edu
Diego Figueiras (PhD)	Advisee/Collaborator	figueirasd1@montclair.edu
Mike Morris (PhD)	Client (MBTI)	[mmorris@themyersbriggs.com](ma

Thank you for your consideration,

John Kulas, Ph.D.
President (and also Janitor), eRg

Minneapolis, MN 55421
jtkulas@ergreports.com
(651) 216-3353