Carleton College
One North College Street
Northfield, Minnesota 55057

Department of Mathematics and Statistics
aloy@carleton.edu

507-222-4360
FAX 507-222-4312

December 19, 2017

Dear Professor Bivand,

We wish to submit an original article entitled "ggplot2 Compatible Quantile-Quantile Plots in R" for consideration by *The R Journal*. We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

In this paper, we provide an overview of the **qqplotr** package which provides a complete implementation of Q-Q plots in the **ggplot2** framework, as well as a wide-ranging discussion on Q-Q plots. We have implemented detrended Q-Q plots for the first time, which have been proven to be more powerful than the traditional Q-Q plot design. (See our article in *The American Statistician* entitled "Variations of Q-Q plots: The power of our eyes!" for our full findings.) Additionally, we have implemented the tail-sensitive simultaneous confidence bands first discussed by Aldoir-Noiman et al. in their *TAS* article entitled "The power to see: A new graphical test of normality." Both of these tools are valuable to R users and of great interest to readers of *The R Journal*.

In addition to discussing new contributions of the package, we discuss points that are not specific to the package:

- In one of our examples, we discuss how jittering can be used to aid the distributional assessment of "discretized" (i.e. rounded) data. We do this in the context of heights, where most respondents rounded to the nearest inch. In this context, using a jittered Q-Q plot better approximates the underlying continuous distribution from which these measurements arise. Since we used the **ggplot2** framework to implement our tools, jittering is easily implemented, but this idea extends beyond our package.
- We briefly discuss two commonly used reference lines for Q-Q plots: the "Q-Q line" that passes through two quantiles (often the 0.25 and the 0.75 quantiles) and the identity line. While the Q-Q line is most commonly used, we highlight that this is not the ideal choice for testing, since the data are used twice: once to fit the reference line, and once for the comparison. To stay true to the testing framework, it is best to specify the theoretical distribution against which you are testing and use the identity line as a visual aid.

We have no conflicts of interest to disclose. Please address all correspondence concerning this manuscript to me at `aloy@carleton.edu`. Thank you for your consideration of this manuscript.

Sincerely,

Adam Loy
with Alexandre Almeida and Heike Hofmann