Capitalized Title Here

by Author One, Author Two, Author Three

Abstract An abstract of less than 150 words.

TODO:

- Determine order of authorship
- Review Q-Q plots and P-P plots, including other arrangements, and what is implemented in other packages
- Write about what the package implements
- Give examples
  - Heike: BRFSS example
- Intro/conclusion
- Abstract

Introduction

Univariate distributional assessment is a common thread throughout statistical analyses during both the exploratory and confirmatory stages. When we begin exploring a new data set we often consider the distribution of individual variables before moving on to explore multivariate relationships. After a model has been fit to a data set, we must assess whether the distributional assumptions made were reasonable, and if they are not we then must understand the impact this has on the conclusions. Graphics provide are arguably the most common way to carry out these univariate assessments. While there are many graphics that can be used for distribution exploration and assessment, probability plotting is one of the most common graphical approaches used.

Probability plotting refers to a family of methods based on the cumulative distribution function (CDF), most notably quantile (Q-Q) plots and probability (P-P) plots (Wilk and Gnanadesikan, 1968). Figure 1 displays two CDFs to illustrate the definition of Q-Q and P-P plots. Let $F_x(q)$ and $F_y(q)$ denote the two CDFs, and let $q_x(p) = F_x^{-1}(p)$ and $q_y(p) = F_y^{-1}(p)$ denote the quantile functions for each distribution. A Q-Q plot is constructed by plotting the $q_y(p)$ against $q_x(p)$ for various p. A P-P plot is constructed by plotting $F_x(q)$ against $F_y(q)$ for various quantiles, q. In practice, the distribution of a sample is often compared to a theoretical distribution. Here, the empirical CDF is used in place of F_y. Regardless of the plot constructed, if the two distributions are the same, then the scattplots will be linear with slope 1 and intercept 0. Additionally, Q-Q plots are invariant to linear transformations, so if two random variables differ by a linear transformation a Q-Q plot showing draws from their distributions will still be linear, but with a different slope and intercept. P-P plots are sensitive to linear transformations.

While the basic form of both the Q-Q and P-P plots is a scatterplot, additional graphical elements are often added to aid in distributional assessment. For Q-Q plots, a reference line is often drawn through the points $(F_x(.25), F_y(.25))$ and $(F_x(.75), F_y(.75))$. For P-P plots a reference line with slope 1 and intercept 0 is used. In both plots, pointwise or simultaneous confidence bands are often added around the reference line to further aid in the visual assessment.

Innovations to Q-Q and P-P plots have also been proposed. Loy et al. (2016) discuss the creation of detrended Q-Q plots, where the y-axis is changed to show the difference $q_y - q_x$. Consequently, the line representing the agreement between the distribution is the x-axis. Loy et al. (2016) find that detrended Q-Q plots are more powerful than other designs, so long at the y-axis limits are set so that the aspect ratio is kept the same as in the traditional Q-Q plot. In reliability and survival analysis, probability plots often refer to a hybrid probability plot, there the CDF of the proposed theoretical distribution is plotted against the empirical order statistics, and transformations are applied to each axis to linearize the CDF (cf. Meeker and Escobar, 1998, chapter 6). This hybrid probability plot is invariant to linear transformations.

ADD EXAMPLES OF THE Q-Q PLOT CONFIGURATIONS FOR CLARITY

Q-Q plots have been implemented in various forms in R, but none provide a complete implementation of the probability plotting framework. Normal quantile plots, where a sample is compared to the standard normal distribution, are implemented using the `qqplot` and `qqline` in **base** graphics (R Core Team, 2012). `qqmath` in **lattice** provides a general framework for Q-Q plots, comparing a sample to any theoretical distribution by specifying the quantile function (Sarkar, 2008). `qqPlot` in the **car** package also allows for the assessment of non-normal distribution and adds pointwise confidence bands based

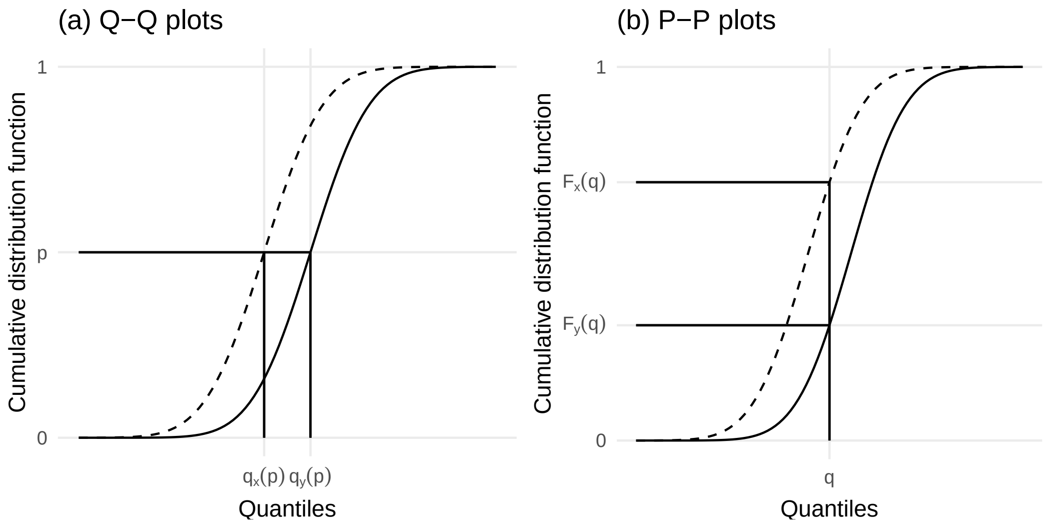

Figure 1: Illustrating what quantities are being plotted for Q-Q and P-P plots.

on the standard errors of the order statistics or the parametric bootstrap (Fox and Weisberg, 2011). **ggplot2** provides `geom_qq` and `geom_qq_line`, enabling the creation of traditional Q-Q plots with a reference line, much like those created using `qqmath`. **qqplotr** extends \pkg{ggplot2 to provide the most complete implementation of probability plotting.

In the remainder of this paper, we introduce the probability plotting framework provided by **qqplotr**. . . FILL THIS IN ONCE OTHER SECTIONS ARE WRITTEN. . .

TODO: FIGURE OUT WHERE TO INTRODUCE TS BANDS (Aldor-Noiman et al., 2013)

Implementing probability plots in the ggplot2 framework

Examples

In this section, we demonstrate the capabilities of the **qqplotr** package.

```
library(qqplotr)
```

Summary

This file is only a basic article template. For full details of *The R Journal* style and information on how to prepare your article for submission, see the Instructions for Authors.

Bibliography

S. Aldor-Noiman, L. D. Brown, A. Buja, W. Rolke, and R. A. Stine. The power to see: A new graphical test of normality. *The American Statistician*, 67(4):249–260, 1 Nov. 2013. [p2]

J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition, 2011. URL `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion`. [p2]

A. Loy, L. Follett, and H. Hofmann. Variations of Q–Q plots: The power of our eyes! *The American Statistician*, 70(2):202–214, 2 Apr. 2016. [p1]

W. Q. Meeker and L. A. Escobar. *Statistical methods for reliability data*. John Wiley & Sons, 1998. [p1]

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org/`. ISBN 3-900051-07-0. [p1]

D. Sarkar. *Lattice: Multivariate Data Visualization with R*. Springer, New York, 2008. URL `http://lmdvr.r-forge.r-project.org`. ISBN 978-0-387-75968-5. [p1]

M. B. Wilk and R. Gnanadesikan. Probability plotting methods for the analysis of data. *Biometrika*, 55 (1):1–17, Mar. 1968. [p1]

Acknowledgements

Mention GSoC here...

Author One
Affiliation
line 1
line 2
author1@work

Author Two
Affiliation
line 1
line 2
author2@work

Author Three
Affiliation
line 1
line 2
author3@work