

Course: Census Data with R

Using the R Package RankingProject to Make Simple Visualizations for Comparing Populations

Developed and presented by Jerzy Wieczorek.

Skill level: Advanced

Duration: 1-2 hours

DESCRIPTION

This course introduces the "RankingProject" package in R, which accompanies "A Primer on Visualizations for Comparing Populations, Including the Issue of Overlapping Confidence Intervals" (Wright, Klein, and Wieczorek, 2018). In comparing a collection of K populations, it is common practice to display K confidence intervals (CIs) for the corresponding population parameters on a single graph. For a pair of CIs that do (or do not) overlap, many viewers find it natural to declare that there is not (or there is) a statistically significant difference between the two corresponding parameters, even though it is well known that this interpretation is not strictly correct.

We will discuss several alternative visualizations designed to help data users avoid this common misinterpretation. CIs for differences from a baseline make the reference population explicit. "Comparison intervals" show a CI for the reference as well as CIs for its difference with other populations. "Shaded columns plots" show the statistical significance of differences directly. Goldstein-Healy adjusted CIs show a confidence level chosen such that overlap (non-overlap) of CIs does indeed imply non-significance (significance) of differences at an "average significance level" across all possible pairwise comparisons. Two-tiered error bars allow us to show several types of CIs at once.

We will justify and recommend use-cases for each of these plots. Finally, we will demonstrate how to produce them in R with the RankingProject package, illustrating its usage on several U.S. Census Bureau datasets with a variety of population types and demographic variables.

WHO SHOULD TAKE THIS COURSE?

Data Analysts, Data Scientists and developers who wish to learn more about how to use Census Data with R to create visualizations.

INSTRUCTOR:

Jerzy Wieczorek is an Assistant Professor of Statistics at Colby College. His research focuses on model selection and assessment, from cross-validation in high-dimensional settings to multiple comparisons-corrected visualization of estimates with uncertainty.