**eyetrackingR: An R Library for Eyetracking Data Analysis**Brock Ferguson (brock@u.northwestern.edu)  
Northwestern University

*Last updated: June 25, 2013*

## Introduction

Psychologists have used eye gaze as a window into cognitive processes for more than a century {Buswell:1922up, Huey:1908tu, Just:1976we, Duchowski:2007uw}, and their measurement has yielded tremendous evolutions in cognitive theory. Likewise, the tools employed to measure eye gaze have evolved and are now capable of answering even more nuanced questions about cognition. Many researchers now employ commercial eyetrackers from companies such as Tobii and SR Research (EyeLink) to record gaze directions and locations. These systems have advantages over traditional methods such as frame-by-frame video coding. For example, they offer greater temporal and spatial resolution, with some current systems sampling gaze direction up to 2000 times per second. However, given the amount of data exported by these systems, comprehensive data analyses can be problematic for psychologists who rely on traditional software to analyse their data (e.g., SPSS, JMP, Excel). Although these packages allow one to model their data, they do not allow for the types of pre-processing and data aggregation required for eyetracking data analysis. Preparing, visualizing, and analysing these data requires specialized methods that can handle the size of eyetracking datasets and the problems unique to their analysis.

In this paper, I introduce *eyetrackerR*, an open source library for R {RAlanguageanden:2012wf} that significantly reduces the time and technical knowledge required to prepare and analyse eyetracking datasets. In section 2, I introduce the basics of R as well as the minimum system and data requirements that must be met to use eyetrackerR effectively. Section 3 reviews methods for summarizing, validating, and visualizing your data. Section 4 reviews methods for preparing your dataset for analysis by selectively removing trials with unacceptable levels of trackloss or non-task directed looking. Finally, section 5 reviews six types of gaze analyses and how they can be implemented in this system. This review will primarily use examples that are familiar to developmental psychologists, however the basic concepts and analysis methods generalize to most common experimental designs employed by psychologists.

## R basics and eyetrackerR requirements

R is a statistical programming language that can be installed on any Mac, Windows, or \*nix system {RAlanguageanden:2012wf}. It is typically used in concert with a Graphical User Interface such as RStudio (http://www.rstudio.com) or RCommander (http://www.rcommander.com). Data is imported into an R workspace environment by reading data files in, for example, the Comma-Separated Values (CSV), Tab-Separated Values (TSV), or Microsoft Excel formats. R’s methods, like the methods of packages or libraries written in R, are called via a text command line interface. Data analyses thus involve importing your data and running a series of commands that augment the data itself (e.g., via transformations, aggregation, or subsetting) or employ statistical methods (e.g., linear models) to estimate parameters and test hypotheses about underlying data models.

eyetrackerR can be downloaded from http://www.github.com/brockf/eyetrackerR, and has a few minimum system requirements. First, R must be installed. Second, the plot\_data() method requires that the ‘ggplot2’, ‘ggthemes’, and ‘reshape2’ packages are installed. Third, it assumes that you have a dataset that includes columns for the participant’s subject code, the Area of Interest (AOI) being gazed at, whether or not the sample was lost to trackloss, the time (ms) relative to some relevant zero point (e.g., trial onset), the sample number, the trial or item name, and a column for each AOI. This dataset should require little if any modification from the data exported by your eyetracker, and should look like the following:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Table 1*  Subset of an example dataset required for eyetrackerR | | | | | | | | |
| SubjID | AgeM | AOI | Trackloss | TimeMS | SampleNum | Trial | Target | Distractor |
| SUB1 | 24.33 | 0 | 1 | 0 | 1 | Monkey | 0 | 0 |
| SUB1 | 24.33 | 0 | 1 | 16.67 | 2 | Monkey | 0 | 0 |
| SUB1 | 24.33 | Target | 0 | 33.34 | 3 | Monkey | 1 | 0 |
| SUB1 | 24.33 | Target | 0 | 50.01 | 4 | Monkey | 1 | 0 |
| SUB1 | 24.33 | Target | 0 | 66.68 | 5 | Monkey | 1 | 0 |
| SUB1 | 24.33 | Target | 0 | 83.35 | 6 | Monkey | 1 | 0 |
| SUB1 | 24.33 | Distractor | 0 | 100.02 | 7 | Monkey | 0 | 1 |
| SUB1 | 24.33 | 0 | 1 | 116.69 | 8 | Monkey | 0 | 0 |
| SUB1 | 24.33 | Distractor | 0 | 133.36 | 9 | Monkey | 0 | 1 |

This dataset also includes an additional column (AgeM) that is a factor that will be of interest to the analysis. Your dataset can include any number of additional factors.

To begin with R, you need only open R, read in your dataset, and source the eyetrackerR library: