

---

# CURRENT STATUS AND PROSPECTS OF R-PACKAGES FOR THE DESIGN OF EXPERIMENTS

---

A PREPRINT

**Emi Tanaka \***

Department of Econometrics and Business Statistics  
Monash University  
Clayton, VIC 3800  
`emi.tanaka@monash.edu`

**Dewi Amaliah**

Department of Econometrics and Business Statistics  
Monash University  
Clayton, VIC 3800

June 6, 2022

## Abstract

Re-running an experiment is generally costly and in some cases impossible due to limited resources, so the design of an experiment plays a critical role in increasing the quality of experimental data. In this article we describe the current state of the R-packages for the design of experiments through exploratory data analysis of package downloads, package metadata, and comparison of characteristics with other topics. We discuss also the software design of widely utilised R packages in the field of experimental design and conclude with discussion of some future prospects for the field.

## 1 Introduction

The critical role of data collection is well captured in the expression “garbage in, garbage out” – in other words, if the collected data is rubbish then no analysis, however complex it may be, can make something out of it. Methods for data collection can be dichotomised by the type of data collected – namely, experimental or observational – or alternatively, categorised as experimental design (including quasi-experimental design) or survey design. This dichotomisation, to a great extent, is seen in the CRAN task views (a volunteered maintained list of R-packages by topic) where R-packages in experimental design are in `ctv{ExperimentalDesign}` and R-packages in survey designs are in `ctv{OfficialStatistics}`. Experimental designs exclusively center on the collection of experimental data where many of the experimental factors tend to be in the control of the experimenter. A subset of experimental designs are segregated into `ctv{ClinicalTrials}`, where the focus is on clinical trials with primary interest in sample size calculations. This paper focuses on the packages in `ctv{ExperimentalDesign}`, henceforth referred to as “DoE packages”.

In the CRAN task view of `ctv{ExperimentalDesign}`, there are 114 R packages for experimental design and analysis of data from experiments. The sheer quantity and variation of experimental designs in the R-packages are arguably unmatched with any other programming languages, e.g. in Python (Rossum 1995), only a handful of libraries that generate design of experiment exist (namely `pyDOE`, `pyDOE2`, `dexpy`, `experimenter` and `GPdoemd`) with limited type of designs. Thus, the study of DoE packages can be revealing into the current status of the field of experimental design.

---

\*[emitanaka.org](http://emitanaka.org)

The paper is organised as follows. Section @ref(data) briefly describes the data source used for the analysis in Section @ref(eda); Section @ref(eda) presents some insights into the state of the current DoE packages by exploratory data analysis of package download data, text descriptions and comparisons with other CRAN task views; Section @ref(design) discuss the interface designs of DoE packages, and we conclude with a discussion in Section @ref(discussion) of future prospects.

## 2 Data

To study the DoE packages, we analyse data using three sources of data as described next.

### 2.1 RStudio CRAN download logs

The Comprehensive R Archive Network (CRAN) is a network of servers located across the world that store mirrored versions of R and R-packages. The most popular network is the RStudio mirror (the default server for those that use the RStudio IDE). The RStudio mirror is also the only server that provides a comprehensive daily download logs of R and R-packages since October 2012. The summary data can be easily accessed with the `cranlogs` package (Csárdi 2019). This paper uses the data from the beginning of 2013 to end of 2021 (a total of 9 years) for the DoE packages.

### 2.2 Package descriptions

All CRAN packages have a title, description, package connections (suggests, dependency and imports) and other meta-information in the DESCRIPTION file. We use the text data from the title and description (as accessed in 2022-05-23) in Section @ref(topics).

### 2.3 CRAN task views

The list of packages in each CRAN task views (as of 2022-05-25) are used to contrast the characteristics of DoE packages in Section @ref(silo).

## 3 Explorative data analysis

All results presented are derived from exploratory data analysis of observational data; consequently, all interpretations are somewhat speculative and may not be indicative of the true state of the field of experimental design. In particular, any analysis over time are confounded by the fact that the nature of users and package management have also changed over the years. It should be noted that some DoE packages may have been archived or removed from the task view over the years so any cross-sectional analysis presented may not reflect the set of all DoE packages at that particular time period (although we assume such incidences are low).

A subset of DoE packages is not primarily about design of experiments but about the analysis of experimental data. A complete delineation of these packages is difficult as many have some functions that can aid decisions or constructions of experimental designs (and any categorisation is prone to our subjective bias) so we opted not to remove any DoE packages in the analysis.

### 3.1 Small, but diverse, set of packages are sufficient for most designs

There have been at least 50 DoE packages since 2013 but most of the downloads are concentrated in just a handful of packages. For example, Figure @ref(fig:plot-lorenz) shows a Lorenz curve (Lorenz 1905) for the total package downloads in 2021 of 113 DoE packages (first released prior to 2021); we can see from Figure @ref(fig:plot-lorenz) that bottom 90% of DoE packages (in terms of total download count in 2021) only share about 30% of total downloads across all DoE packages – in another words, 70% of the total downloads are due to 11 packages (10% of the DoE packages).

If we consider package downloads as a measure of “wealth”, then we can consider using the Gini index (Gini 1921) as a measure of download inequality across packages. The ratio of the red region over the total colored regions in Figure @ref(fig:plot-lorenz) corresponds to the Gini index for 2021. A Gini index of 0% indicates equality in downloads across packages while a value of 100% indicates maximal inequality (all downloads are due to one package). In Figure @ref(fig:download-share), we see that the distributions of the downloads

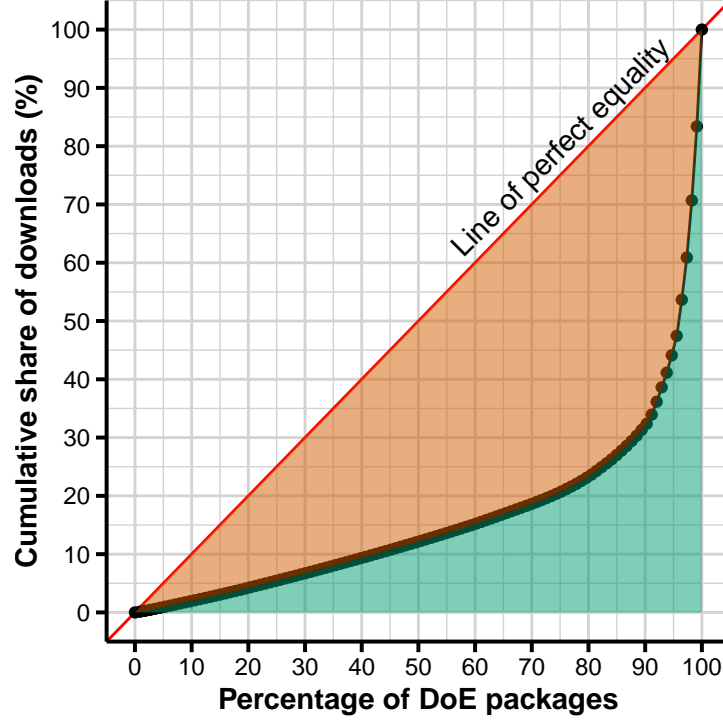


Figure 1: Lorenz curve of the total download count for DoE packages in 2021. The red line corresponds to the line of perfect equality.

each year have a heavy right tail with the Gini index ranging from 32.5% to 68.3% across the years 2013 to 2021, indicating that there is a high level of inequality of downloads across packages, particularly with more pronounced inequality in the last 5 years.

While in absolute terms the Gini index is high for `ctv{ExperimentalDesign}`, the inequality is not as severe as other CRAN task views as shown in Figure @ref(fig:fig-gini-all-ctvs). We can see in Figure @ref(fig:fig-gini-all-ctvs) that the Gini index is generally increasing for DoE packages (as is generally the case for other CRAN task views as shown in Figure S1 in the Supplementary Materials).

From Figures @ref(fig:download-share) and @ref(fig:fig-gini-all-ctvs), we could draw some indicative states of the field of experimental designs with possible counterfactual interpretations:

- The increase in the number of packages that are not highly downloaded may mean that ***there are more packages to construct niche experimental designs***. Some examples of these packages include `qtlDesign`, `PwrGSD` and `Crossover` made for QTL experiments, group sequential designs and crossover trials, respectively. These packages would naturally have a smaller number of potential users. Counterfactual to this, the increase could be due to other external factors, such as an increase in a number of skilled contributors, a change in CRAN policy or management to add packages (either to CRAN and/or task view), and/or that new packages are still yet to amass users. While there is an argument that low download counts are due to the low utility and/or quality of the packages, packages in CRAN task views are selected by expert maintainers; thus we can reasonably assume that any package listed in CRAN task view are of a decent utility and quality.
- If the downloads are reflective of the experimental designs used in practice, ***small set of packages appear to be sufficient for most to construct the full set of designs of experiments needed in practice***. Packages of course evolve and the top downloaded packages have had regular updates that may have broadened its scope.
- While small subset of DoE packages are most frequently used, none of these packages are dominant in the field (judging from comparisons with other CRAN task views). This suggests that there are ***diverse approaches to designing experiments***. This observation doesn't take into account

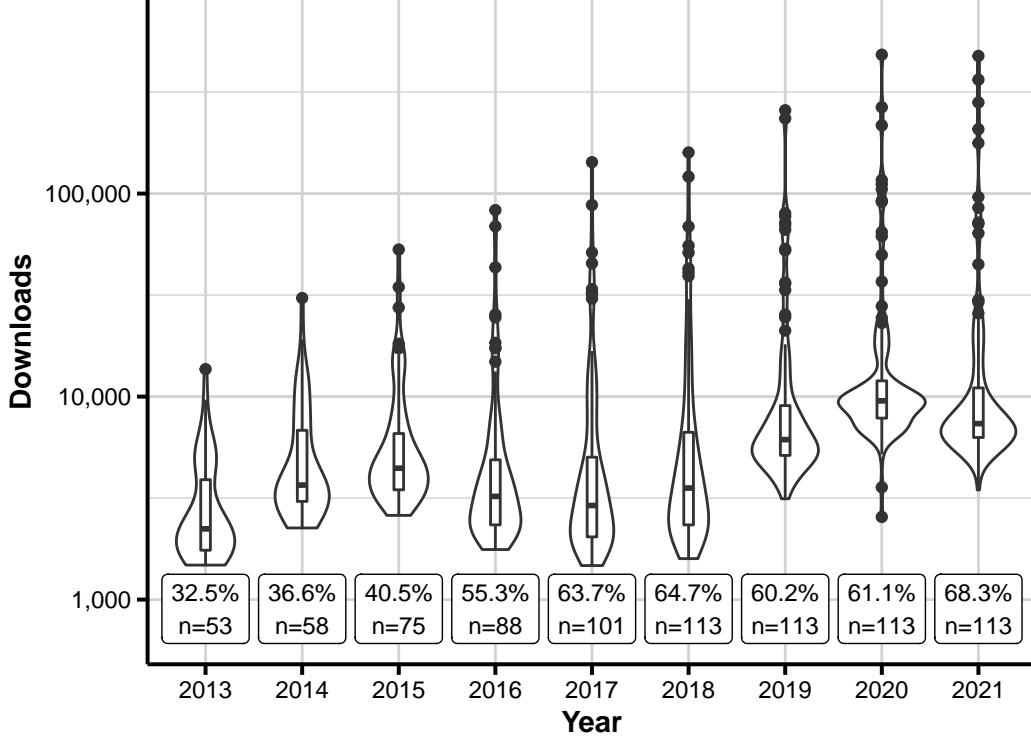


Figure 2: The distribution of the number of downloads for each DoE package by year. Packages were removed in any year if it was released in that year or later so that each download count is for the full year. The label on the bottom of the plot shows the Gini index for downloads and the number of packages with full year of download count in the corresponding year. In the last 5 years, the Gini index is consistent above 60% each year indicating that most downloads are due to a relatively small number of packages.

other approaches to generating experimental designs, like the proprietary GUI software, CycDesignN, which may be widely used.

### 3.2 There is a lack of adaptation of new or innovative designs

We can see in Figure @ref(fig:rank-over-time) that most of the top 10 ranking packages have been in the top 10 for the last 9 years with `lhs` steadily climbing up the ranks in the last few years.

Figure @ref(fig:established-packages) shows a moderate negative correlation between the first release date and the (log of) total download counts of DoE packages in any given year from 2013 to 2021. This suggest that in general, a package released earlier are more likely to be used today (possibly for legacy reasons or the general inertia for adoption of new packages). We can also see in Figure @ref(fig:established-packages) the most downloaded packages were generally released in 2004 to 2010.

The lack of change in the top ranking packages are indicative that there has been generally little need for new or innovative experimental design needs for the mass. We do however also see that the the top downloaded packages generally have more updates so it is possible that the packages have improved or broadened the scope of its usage.

### 3.3 Computer experiments and optimal designs are of interest

Figure @ref(fig:wordcloud-over-time) shows some common purposes of DoE packages, based on bigrams in the package title and description. A bigram is only shown as the frequency of single word was not insightful and there were not many trigrams common across packages. Not surprisingly, the bigram “experimental design” was the most common. More interestingly, “optimal design” appeared across different packages (indicated by the size of the word in Figure @ref(fig:wordcloud-over-time)), and the bigrams “latin hypercube” and “computer experiment” used across a few packages that are downloaded frequently (indicated by the color of

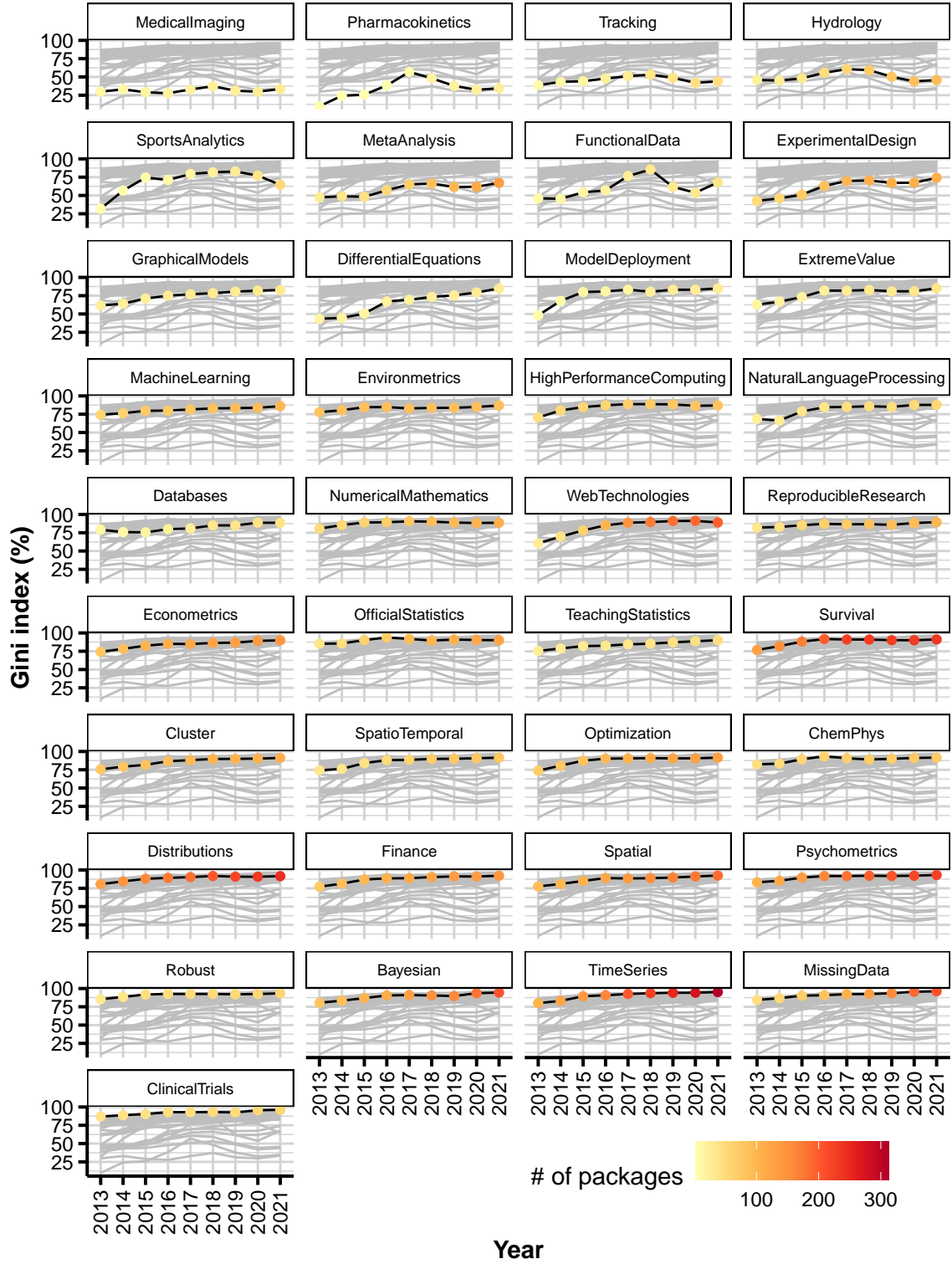


Figure 3: The points show the Gini index of the download counts by year faceted by CRAN task view with the color showing the number of packages. The grey line shows the distribution of the Gini index across years for all other CRAN task views. The facets are ordered by increasing value of the Gini index in 2021.

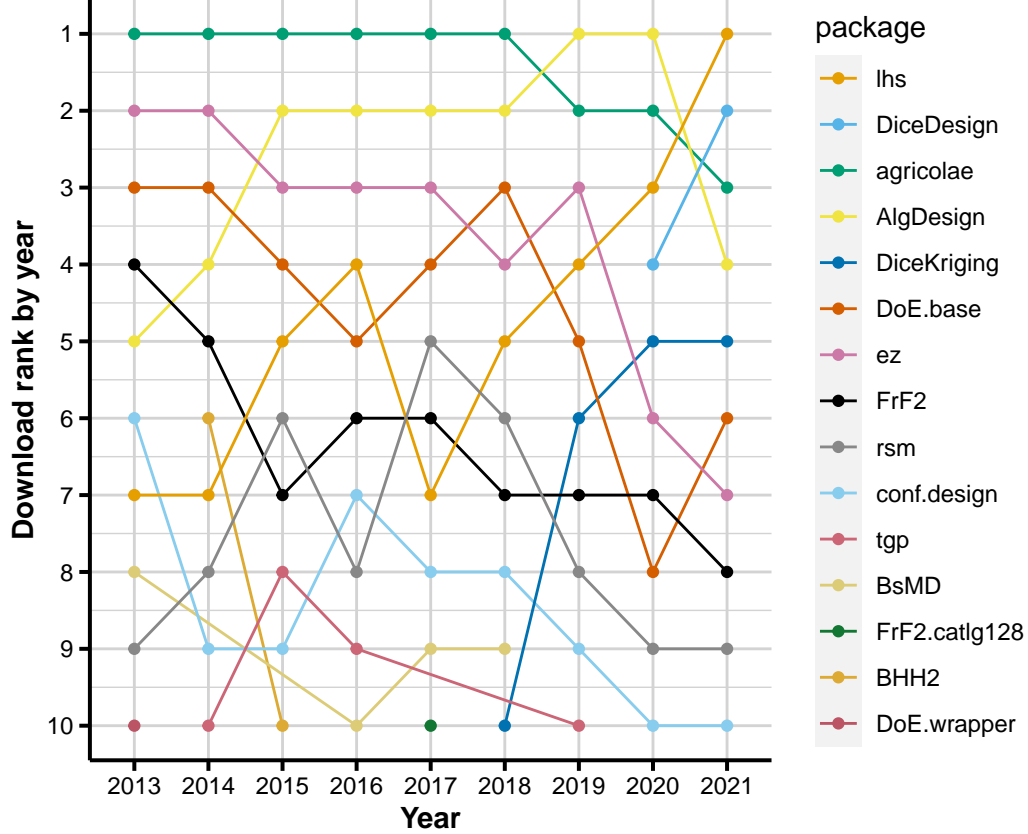


Figure 4: The plot shows the rank of top 10 downloaded packages by year.

the word in Figure @ref(fig:wordcloud-over-time)). Although there exists a separate `ctv{ClinicalTrials}` task view, the DoE packages clearly include some packages that are of interest to clinical trials as shown by the size of the bigram “clinical trial” in Figure @ref(fig:wordcloud-over-time).

### 3.4 Development of experimental designs occur in silos

Figure @ref(fig:ctv-summ-plot) shows that the `ctv{ExperimentalDesign}` have the lowest average number of contributors among all CRAN task views. In addition, we can also see in Figure @ref(fig:ctv-summ-plot) that the `ctv{ExperimentalDesign}` task view has one of the least intra-connectivity (the percentage of packages that make use of other packages within the same task view). The full connection of DoE packages with other DoE packages is shown in Figure @ref(fig:plot-doe-network). These observations suggest that the field of experimental design is one of the least collaborative field and package development generally occur in silos.

## 4 Interface design

In software design, there are two interface designs to consider: user interface (UI) and application programming interface (API). UI is concerned with the usage of the software by the user, while the API is concerned with how different programs interact and is predominately of the interest to the developer. In this section, we discuss these interface designs broadly, with examples illustrated from the top downloaded packages (as shown in Figure @ref(fig:rank-over-time)). We exclude `ez` and `DiceKriging` from the discussion as the former is predominately visualisation of experimental data and latter is about the analysis of computer experiments in addition to belonging to the same suite of packages as `DiceDesign`.

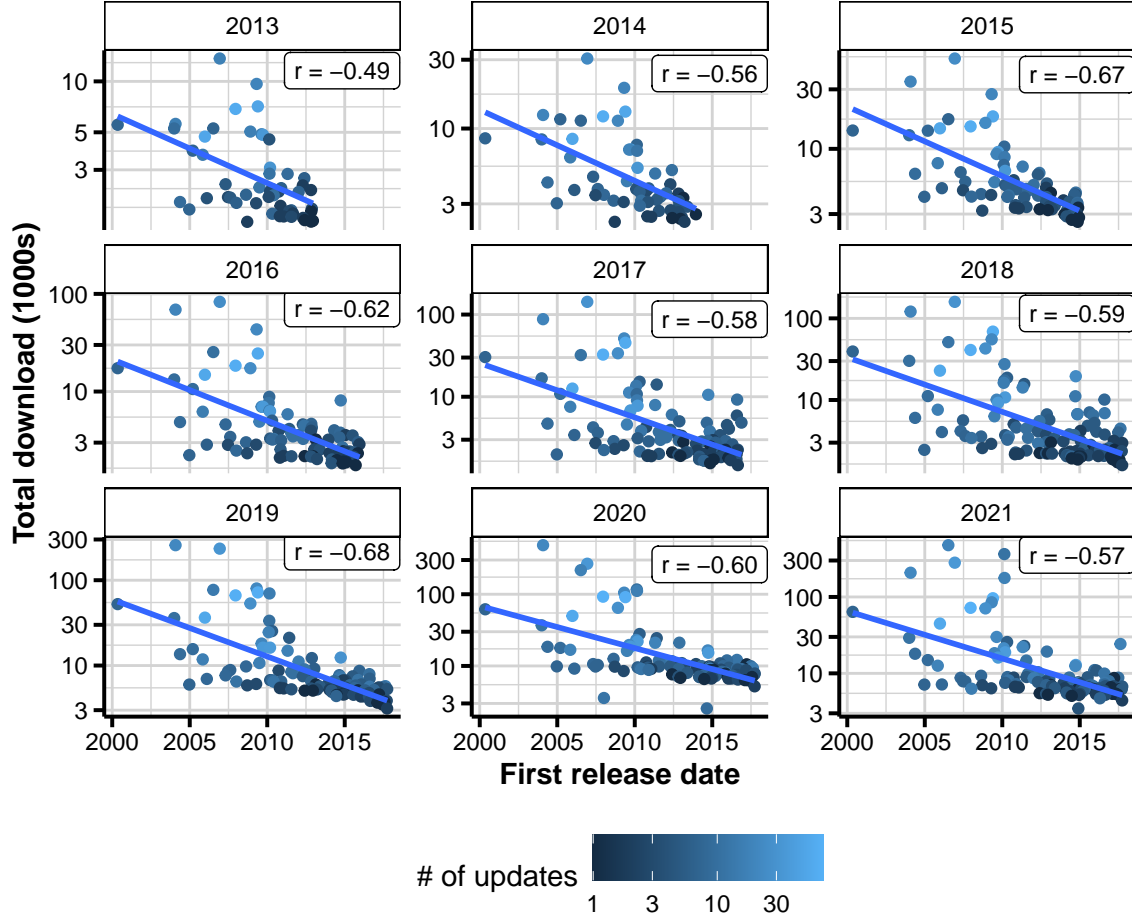


Figure 5: The above figure shows the total download (in log scale) of a package in the corresponding year against the first release date of the package. The blue line corresponds to the least squares fit of a simple linear regression model. The label in the right-hand upper corner shows the sample correlation coefficient between the first release date and the log, with base 10, of the total download count. The high-leverage point on the far left belongs to ‘conf.design’, authored by one of the earlier contributors to R.

#### 4.1 The case of specialised designs

In this section, we present some examples the user interface of widely used specialised designs, namely, computer experiments, factorial experiments, response surface designs and sequential designs.

Computer experiments, as discussed in Section @ref(topics), which generally involve space filling designs have been on the rise. The exemplar for this type of designs are `lhs` and `DiceDesign`. For the `lhs` package, functions like `randomLHS()`, `optimumLHS()`, and `maximinLHS()`

Factorial experiments offer a challenge in the construction and allocation of the treatment factors. This is reflected in a number of packages that specifically addresses this challenge. These include `DoE.base` (Grömping 2018), `FrF2` (Grömping 2014) for fractional 2-level factorial designs, `conf.design` (Venables 2013), `FrF2.catlg128` (Grömping 2022) and `BHH2` (Barrios 2016).

Response surface design, as illustrated by `rsm` (Lenth 2009), is a type of experiment that involve variables with continuous measures.

Finally, another specialised design is sequential design (also called adaptive sampling), which is best represented by `tgp` (Gramacy and Taddy 2010) – these require prior information, which are used to inform the next experimental design. Follow-up experiments, which can also be classified as a sequential design, is implemented by `BsMD` (Barrios 2020).



Figure 6: The above figure shows the word cloud of bigrams from the title and descriptions of the CRAN packages. The size shows how often the bigram appears across the DoE packages and the color are relative to the total download count in 2021 of the packages that contain the bigram.

## 4.2 The case of menu functions

The **agricolae** package (de Mendiburu 2021) was the most downloaded DoE packages in 2013 to 2018 (Figure @ref(fig:rank-over-time)) and possibly prior to 2013 (data unavailable). This package is the prime example of constructing designs based on menu functions, e.g. `design.crd()`, `design.rcbd()` and `design.split()` construct a completely randomised design, randomised complete block design and split-plot design, respectively. Users typically supply the treatment labels and the number of replications as argument to these functions.

## 4.3 The case of general designs

**AlgDesign** (Wheeler 2022)

**DoE.wrapper** (Grömping 2020)

## 5 Discussion

Volunteer maintained – some experimental design packages may not be included in the CRAN task view.

## 6 Acknowledgement

This paper uses the **targets** framework (Landau 2021) for reproducibility, **knitr** (Xie 2015) and **rmarkdown** (Xie, Allaire, and Grolemund 2018) for creating reproducible documents, **ggplot2** (Wickham 2016) for



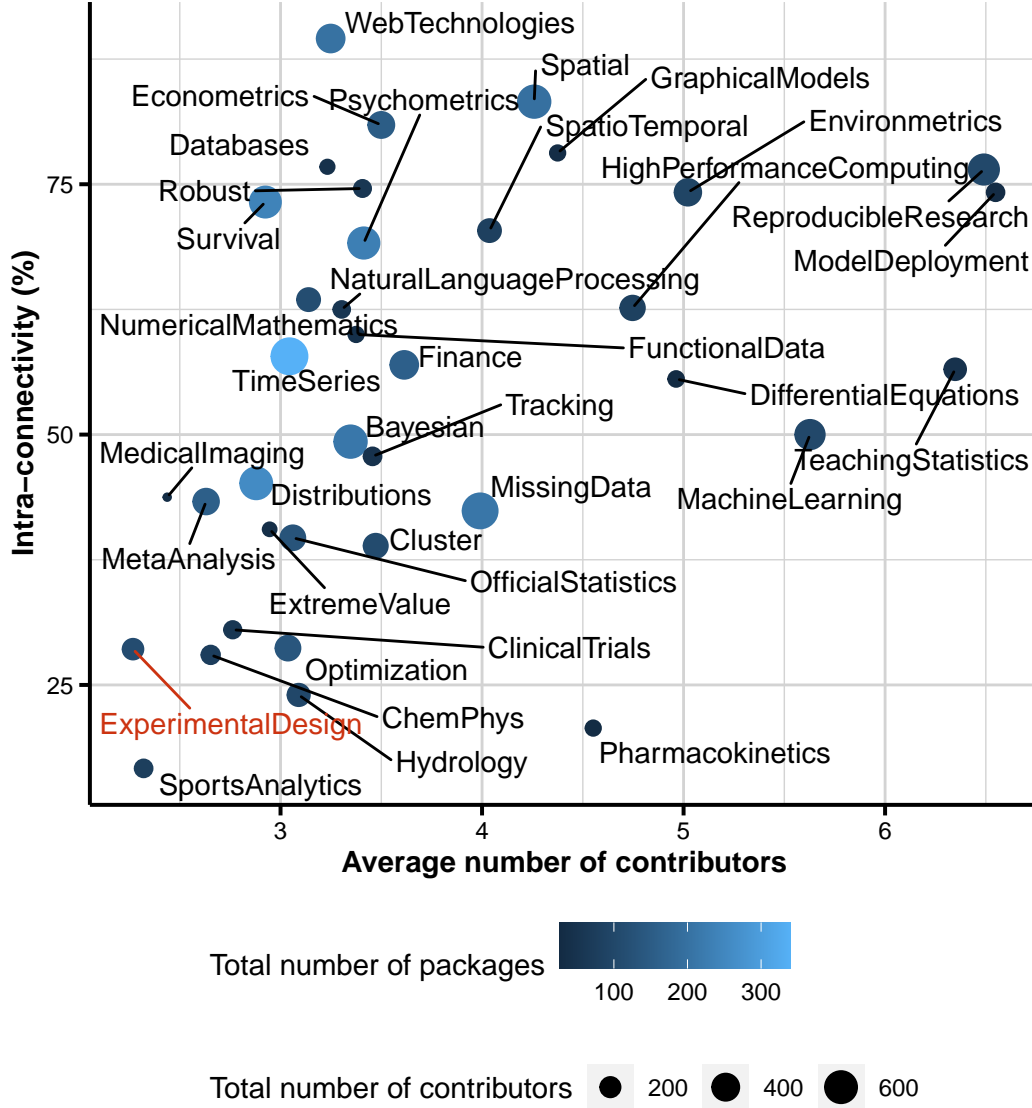


Figure 7: The above figure is a scatterplot of the intra-connectivity (the percentage of packages that depends, suggest or imports at least one other package within the same task view) and the average number of contributors for each CRAN task view. A low intra-connectivity suggests that development within the topic mostly occur in silos whilst high intra-connectivity suggests that there are more interactions within the topic. The color shows the number of packages, the size of the point corresponds to the total number of contributors, and the text labels show the CRAN task view name. The label of ExperimentalDesign task view is colored red. The task views in the bottom-left corner are topics that are more indicative of contributors working in silos. The actual numerical values are show in Table S1 in the Supplementary Material.

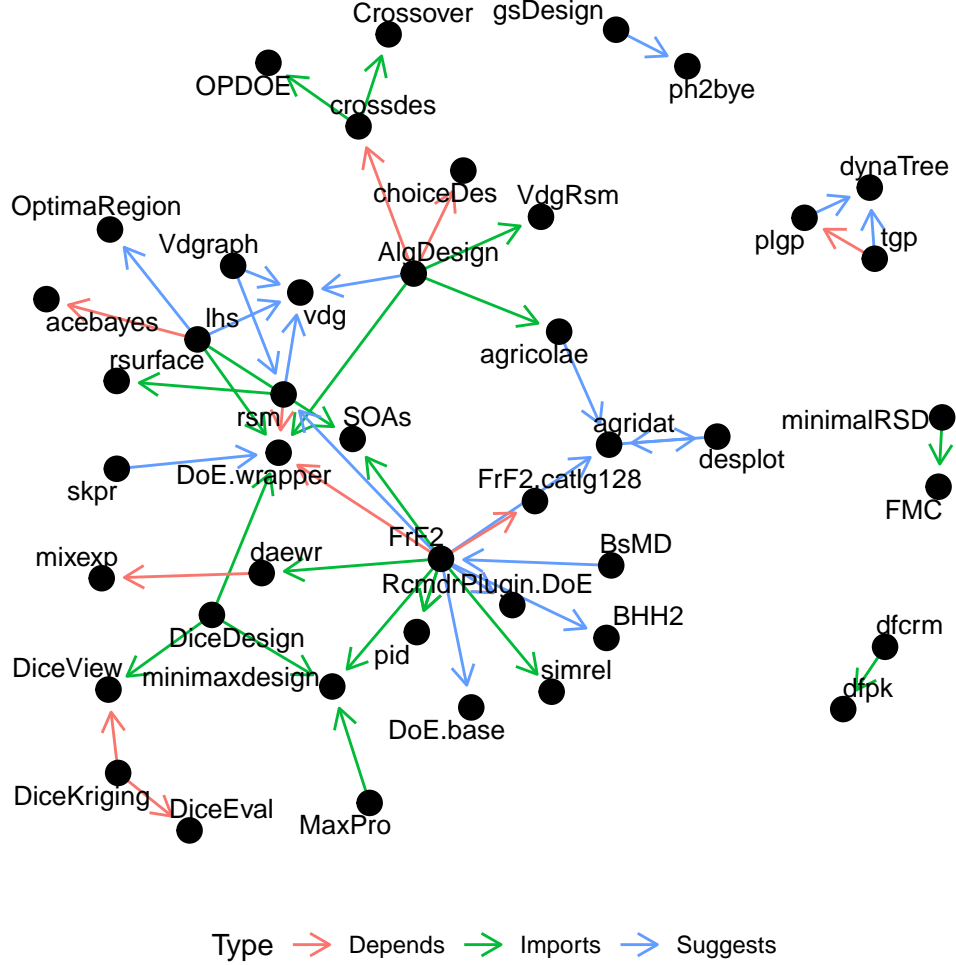


Figure 8: Package connections (depends, suggests and imports) within the DoE packages. DoE packages that do not depend, suggest or import another DoE package is not shown.

visualisation and `kableExtra` (Zhu 2021) for customising the table. All code to reproduce this paper is found at <https://github.com/emitanaka/paper-DoE-review>.

## References

- Barrios, Ernesto. 2016. *Bhh2: Useful Functions for Box, Hunter and Hunter II*. <https://CRAN.R-project.org/package=BHH2>.
- . 2020. *BsMD: Bayes Screening and Model Discrimination*. <https://CRAN.R-project.org/package=BsMD>.
- Csárdi, Gábor. 2019. *Cranlogs: Download Logs from the 'RStudio' 'CRAN' Mirror*. <https://CRAN.R-project.org/package=cranlogs>.
- de Mendiburu, Felipe. 2021. *Agricolae: Statistical Procedures for Agricultural Research*. <https://CRAN.R-project.org/package=agricolae>.
- Gini, Corrado. 1921. “Measurement of Inequality of Incomes.” *The Economic Journal* 31 (121): 124–26. <https://doi.org/10.2307/2223319>.
- Gramacy, Robert B., and Matthew Taddy. 2010. “Categorical Inputs, Sensitivity Analysis, Optimization and Importance Tempering with `tgp` Version 2, an R Package for Treed Gaussian Process Models.” *Journal of Statistical Software* 33 (6): 1–48. <https://doi.org/10.18637/jss.v033.i06>.
- Grömping, Ulrike. 2014. “R Package `FrF2` for Creating and Analyzing Fractional Factorial 2-Level Designs.” *Journal of Statistical Software* 56 (1): 1–56. <https://www.jstatsoft.org/v56/i01/>.

- . 2018. “R Package DoE.base for Factorial Experiments.” *Journal of Statistical Software* 85 (5): 1–41. <https://doi.org/10.18637/jss.v085.i05>.
- . 2020. *DoE.wrapper: Wrapper Package for Design of Experiments Functionality*. <https://CRAN.R-project.org/package=DoE.wrapper>.
- . 2022. *FrF2.catlg128: Catalogues of Resolution IV 128 Run 2-Level Fractional Factorials up to 33 Factors That Do Have 5-Letter Words*. <https://CRAN.R-project.org/package=FrF2.catlg128>.
- Landau, William Michael. 2021. “The Targets r Package: A Dynamic Make-Like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing.” *Journal of Open Source Software* 6 (57): 2959. <https://doi.org/10.21105/joss.02959>.
- Lenth, Russell V. 2009. “Response-Surface Methods in R, Using rsm.” *Journal of Statistical Software* 32 (7): 1–17. <https://doi.org/10.18637/jss.v032.i07>.
- Lorenz, M O. 1905. “Methods of Measuring the Concentration of Wealth.” *Publications of the American Statistical Association* 9 (70): 209–19. <https://doi.org/10.2307/2276207>.
- Rossum, G. van. 1995. “Python Tutorial.” CS-R9526. Amsterdam: Centrum voor Wiskunde en Informatica (CWI).
- Venables, Bill. 2013. *Conf.design: Construction of Factorial Designs*. <https://CRAN.R-project.org/package=conf.design>.
- Wheeler, Bob. 2022. *AlgDesign: Algorithmic Experimental Design*. <https://CRAN.R-project.org/package=AlgDesign>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- Xie, Yihui, J. J. Allaire, and Garrett Grolemond. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.