

---

## TP : Introduction à R

### Objectif :

- prendre en main les concepts de base de R, RStudio et de Rmd
- réaliser des ANOVA

## 1 Introduction

R<sup>1</sup> est un logiciel de statistiques gratuit et open-source, disponible sous Linux, Windows et OS X.

Pour lancer R, tapez **R** en ligne de commande. Il est également possible d'écrire des scripts et de les interpréter avec R. Pour obtenir de l'aide en ligne sur une commande, il suffit de la faire précéder du symbole `?`. Les flèches haut et bas permettent de naviguer dans l'historique de commandes. Les commentaires sont précédés par le symbole `#`. La touche `tab` peut être utilisée pour la complétion des commandes et des noms de variables. Il est recommandé de travailler avec l'IDE RStudio et d'utiliser la syntaxe Markdown (fichiers avec l'extension Rmd).

## 2 Réplication des analyses statistiques de l'article ForceEdge

L'objectif est de reproduire une partie des analyses statistiques détaillées dans l'article suivant :

*Axel Antoine, Sylvain Malacria, and Géry Casiez. 2017. ForceEdge: Controlling Autoscroll on Both Desktop and Mobile Computers Using the Force<sup>2</sup>. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 3281-3292.*

Nous allons nous concentrer uniquement sur la première expérience. Commencez par lire le protocole expérimental.

**Q 1.** Ouvrez RStudio et créez un nouveau fichier Rmd.

### 2.1 Chargement des données

**Q 2.** Utilisez la commande `read.table`<sup>3</sup> pour charger les données du fichier `ForceEdgeStudy1.csv`<sup>4</sup>.

### 2.2 Filtrage et agrégation des données

**Q 3.** En utilisant les fonctionnalités du package `dplyr`<sup>5</sup>, conservez uniquement les essais pour lesquels il n'y a pas eu d'erreur (voir les exemples du cours<sup>6</sup>).

**Q 4.** Regroupez maintenant les données par **task**, **technique** et **distance** et calculez pour chaque sous-groupe le temps moyen ainsi que les bornes inférieures et supérieures des intervalles de confiance à 95% en utilisant la commande `ci`<sup>7</sup> du package `gmodels`.

---

1. <https://www.r-project.org>

2. [https://gery.casiez.net/IHMA/TP\\_R/3025453.3025605.pdf](https://gery.casiez.net/IHMA/TP_R/3025453.3025605.pdf)

3. <https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/read.table>

4. [https://gery.casiez.net/IHMA/TP\\_R/ForceEdgeStudy1.csv](https://gery.casiez.net/IHMA/TP_R/ForceEdgeStudy1.csv)

5. <https://www.rdocumentation.org/packages/dplyr/versions/0.7.8>

6. [https://gery.casiez.net/IHMA/TP\\_R/CoursNIHMStatsExemples.zip](https://gery.casiez.net/IHMA/TP_R/CoursNIHMStatsExemples.zip)

7. <https://www.rdocumentation.org/packages/gmodels/versions/2.18.1/topics/ci>

## 2.3 Visualisation

**Q 5.** A partir des données calculées précédemment, réalisez un graphique similaire à celui visible figure 4 (gauche) de l'article. Pour cela, utilisez le package `plotly`<sup>8</sup>. Il est également possible d'utiliser le package `ggplot2`<sup>9</sup> et en particulier `geom_line` et `geom_errorbar`.

## 2.4 ANOVA sur le temps

**Q 6.** Dans un premier temps l'objectif est de réaliser une ANOVA pour constater l'effet d'apprentissage (différence significative entre le bloc 1 et les autres). Utilisez la commande `ezANOVA`. Reportez-vous aux exemples du cours.

**Q 7.** L'objectif est maintenant de supprimer les données du premier bloc et relancer l'ANOVA pour analyser les autres variables indépendantes.

## 2.5 Friedman sur les questionnaires NASA-TLX

**Q 8.** L'objectif est ici de refaire les analyses de Friedman sur l'item frustration. Les données sont disponibles ici<sup>10</sup>. Reportez-vous aux exemples du cours.

---

8. <https://plotly-r.com/>

9. <https://ggplot2.tidyverse.org/reference/>

10. [https://gery.casiez.net/IHMA/TP\\_R/ForceEdgeStudy1NasaTLX.csv](https://gery.casiez.net/IHMA/TP_R/ForceEdgeStudy1NasaTLX.csv)