Haguenau

JcB

14/10/2015

# Objectifs

* savoir utiliser un logiciel de statistiques
* savoir collecter correctement des données (tableur)
* transmettre les données au logiciel
* appliquer une démarche statistique

# Organiser son travail

* Démarrer RStudio
* Créer un nouveau \_Projet\_\_ dans un nouveau répertoire (directory): File/New project -> New directory
* Créer un sous répertoire **Data** qui servira à stocker les données
* Créer un sous répertoire **Cours\_Stat\_2015**

# Savoir utiliser un logiciel de statistiques

* Utilisation de **R** (chercher [*CRAN*](https://cran.r-project.org/) The Comprehensive R Archive Network).
* c'est à la fois un language de programmation (on peut écrire ses propres routines) et un logiciel statistique.
* **R** est *libre*, *gratuit*, *multiplateforme*, *complet*, *évolutif* grâce à une énorme bibliothèque de fonctions appelées **Pakages** (environ 6000 à ce jour).
* On peut l'utiliser nativement ou par l'intermédiaire d'un IDE appelé **RStudio**.
* *RStudio* utilise le concept de **recherche reproductible** et permet de mettre en place une chaîne de production allant de la saisie des données à la production d'un document (mémoire, thèse, etc.) publiable.

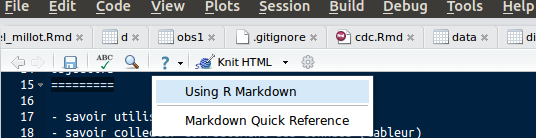
## Caractéristiques de RStudio

### 4 fenêtres

* Ouvrir une nouvelle page de travail: file -> New File -> R Markdown...
* **Enregistrer** le document dans le dossier **Cours\_Stat\_2015**

### RMarkdown

Un [traitement de texte simple](http://rmarkdown.rstudio.com/) pour prendre des **notes**



### les "Chunks"

Aujourd'hui, je crée mon premier programme R en fusionnant mon texte, mes calculs et graphiques dans le même document grace aux *chunks*:

print("Hello, R")

## [1] "Hello, R"

a <- 2 + 2  
a

## [1] 4

Il semble que 2 + 2 fassent 4

### Prouire un document

A partir de ce document écrit avec *markdown* je peut produire directement:

* un document **Html** pour un navigateur (site internet, blog...)
* un document **Pdf** non modifiable à distribuer
* document **Word** ou **Libre Office** modifiable
* un \_diaporama\_\_ à projeter
* un fichier des **graphiques** crées pouvant ^etre copiés/collés dans n'importe quel document.

# La démarche statistique

## 1. Collecte de données

et leur mise en forme pour être exploitées, c'est 80% du travail.

## 2. Statistiques descriptives

Nous allons générer des statistiques qui résument les données de façon concise, et d'évaluer les différents des moyens pour visualiser les données.

## 3. L'analyse exploratoire des données

Nous allons rechercher des modèles, les différences, et d'autres caractéritiques qui répondent aux questions nous sommes intéressés à. Dans le même temps, nous allons vérifier les incohérences et identifier limitations.

## 4. estimation

Nous allons utiliser les données à partir d'un échantillon pour estimer les caractéristiques de la population générale.

## 5. Tests d'hypothèses

Où l'on voit les effets apparents, comme une différence entre deux groupes, nous évaluerons si l'effet pourrait être dû au hasard.

# La collecte des données

Application: analyse des friandises contenues dans un paquet de [M&M's](http://fr.wikipedia.org/wiki/M%26M%27s) fabriquées à Haguenau.

