

Présentation R

Arthur Lemoine, François Somville, Alexandre Antippas

Janvier 2019

Plan

- ▶ Introduction
- ▶ Exploration du Dataset
- ▶ Statistiques descriptives
- ▶ Analyse des composantes principales (ACP)
- ▶ Clustering
- ▶ Bonus

Introduction

Le dataset contient toutes les factures émises par un magasin en ligne basé en Angleterre. Le magasin vend des cadeaux uniques pour toutes les occasions.

Exploration du Dataset

Nombre d'enregistrements

```
## [1] 541909
```

Nombre de variables

```
## [1] 8
```

Exploration du Dataset

Les différentes variables

```
## 'data.frame':    541909 obs. of  8 variables:
## $ InvoiceNo   : Factor w/ 25900 levels "536365","536366",...
## $ StockCode  : Factor w/ 4070 levels "10002","10080",...
## $ Description: Factor w/ 4224 levels "", " 4 PURPLE FLOO
## $ Quantity   : int   6 6 8 6 6 2 6 6 6 32 ...
## $ InvoiceDate: Factor w/ 23260 levels "1/02/11 08:23",...
## $ UnitPrice  : num    2.55 3.39 2.75 3.39 3.39 7.65 4.25
## $ CustomerID : int    17850 17850 17850 17850 17850 17850
## $ Country    : Factor w/ 38 levels "Australia","Austria"
```