

TP_Sirene

Benoît Blanc

10/10/2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

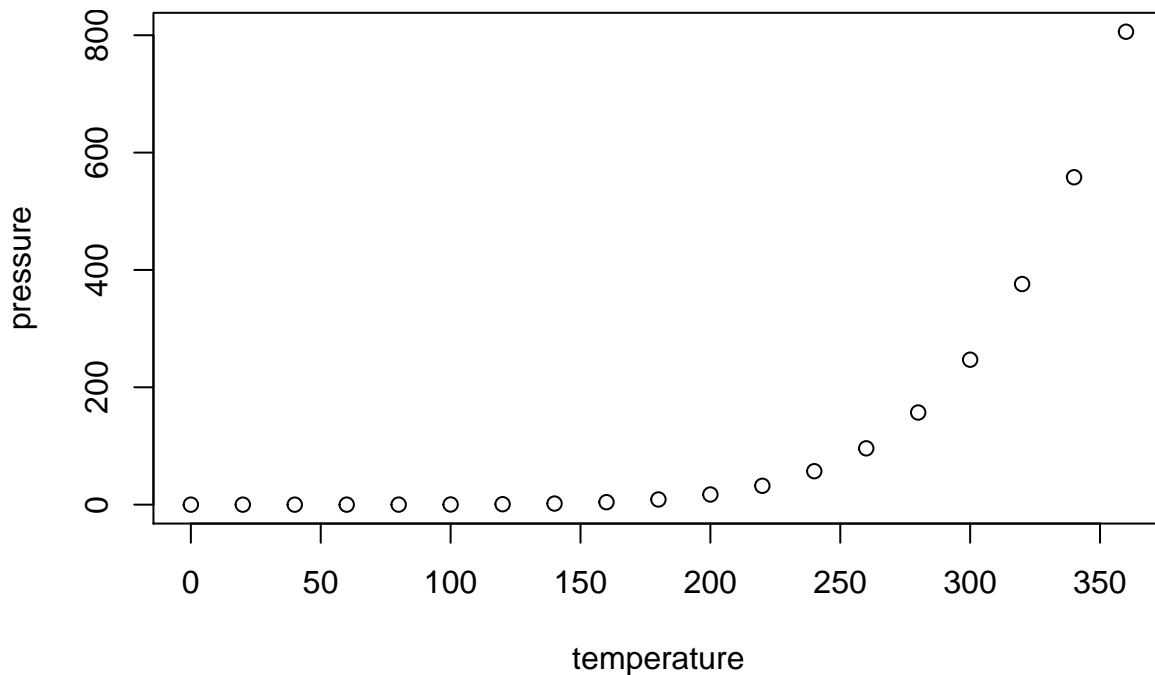
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

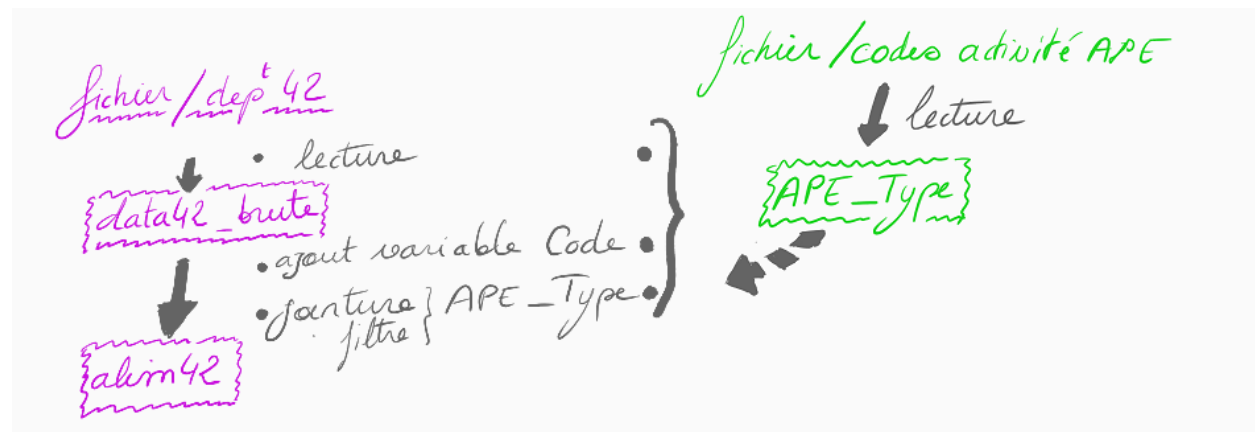
Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Mise en place



Installations, téléchargements, premiers tests sur le département 42

Téléchargez les données du département 42, `geo_siret_42.csv` dans ce répertoire et dézippez le dossier sur votre machine.

Téléchargez la table qui renseigne les codes correspondant à l'Activité Principale de l'Etablissement (APE) `APE_Type.csv`

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(readr)
```

Lecture de tableaux de données

Depuis RStudio, créez un projet qui comprendra l'ensemble des données et documents nécessaires à réaliser l'ensemble des traitements qui vous seront demandés pour ce TP.

Créez le data.frame `data42` en lisant la table `geo-siret_42.csv`.

```
data42 <- read.csv("data/geo_siret_42.csv")
```

Créez l'objet `APE_Type` en lisant le fichier relatif aux codes d'APE.

```
APE_Type <- read.csv("data/APE_Type.csv")
```

Code et types d'activités => commerces alimentaires

```
library(stringr)
```

Combien d'entreprises ont un nom (`enseigne1Etablissement`) qui comprend le terme "BOULANGERIE"?

```
boulangeries <- filter(data42, str_detect(enseigne1Etablissement, "BOULANGERIE"))
```

Il y a 84 entreprises qui comprennent le terme “BOULANGERIE” dans leur nom.

Ajoutez une variable Code à votre table en ne conservant que les quatre premiers caractères de la variable activitePrincipaleEtablissement (cela correspond à un pattern “^...”, à savoir le début de chaîne de caractère suivi de quatre caractères quelconques -cf ce billet de blog sur les expressions régulières-).

```
data42 <- mutate(data42, Code=as.numeric(str_extract(activitePrincipaleEtablissement, "^...")))
```

Filtrez les lignes de data42 pour ne retenir que celles pour lesquelles l’APE correspond aux commerces “alimentaires” -alimentation, boisson, restaurant, bar- (voir la liste contenue dans le fichier APE_Type).

Stockez le résultat de ces opérations dans un objet alim42.

```
alim42 <- filter(data42, Code=="47.1" | Code=="47.2" | Code=="56.1" | Code=="56.2" | Code=="56.3")
```

Réalisez une jointure entre data42_alim (variable codeAPE) et APE_Type (variable Code), de manière à compléter alim42 avec les types de commerces (variables Type et TypeAbreg).

```
alim42 <- left_join(alim42, APE_Type, by="Code")
```

Résumé, classement

Quelles sont les 3 communes de votre base de données qui comptent le plus de magasins alimentaires?

Les trois communes comptant le plus de commerces alimentaires sont

```
alim42 %>%
  group_by(codeCommuneEtablissement, libelleCommuneEtablissement) %>%
  summarise(nb_commerces_alim=n()) %>%
  arrange(desc(nb_commerces_alim)) %>%
  head(3)
```

```
## # A tibble: 3 x 3
## # Groups:   codeCommuneEtablissement [3]
##   codeCommuneEtablissement libelleCommuneEtablissement nb_commerces_alim
##               <int> <fct>                        <int>
## 1             42218 SAINT ETIENNE                      3419
## 2             42187 ROANNE                             802
## 3             42207 SAINT CHAMOND                      452
```

Pour les communes qui ne comptent qu’un seul commerce “alimentaire”, de quel type est-il, le plus fréquemment?

```
onlyonecommerce <- group_by(alim42, codeCommuneEtablissement, libelleCommuneEtablissement) %>%
  summarise(nb_commerces_alim=n()) %>%
  filter(nb_commerces_alim==1)
```

Les communes ne comptant qu’un seul commerce “alimentaire” sont au nombre de 15. Avec en détail : APINAC, BOYER, CEZAY, LA CHAPELLE EN LAFAYE, GREZIEUX LE FROMENTAL, JARNOSSE, PINAY, SAINTE AGATHE EN DONZY, SAINT LAURENT ROCHEFORT, SAINT PIERRE LA NOAILLE, SAINT PRIEST LA VETRE, SAINT SIXTE, TARTARAS, URBISE, VIRICELLES

```
typeonecommerce <- filter(alim42, codeCommuneEtablissement %in% onlyonecommerce$codeCommuneEtablissement)
  group_by(Type) %>%
  summarise(nb_types=n()) %>%
  arrange(desc(nb_types)) %>%
  filter(nb_types == max(nb_types))
```

Lorsqu'il y a un seul commerce alimentaire dans une commune, celui-ci est de type Restaurants et services de restauration mobile. Quelles communes de plus de 100 commerces comptent au moins 10 commerces de type "viande"?