

19/01/2019

1

Introduction à R : créer des graphiques à partir d'un fichier de données

Logiciel de calcul statistique et création de graphique

M. PERES LEBLANC



Télécharger le logiciel



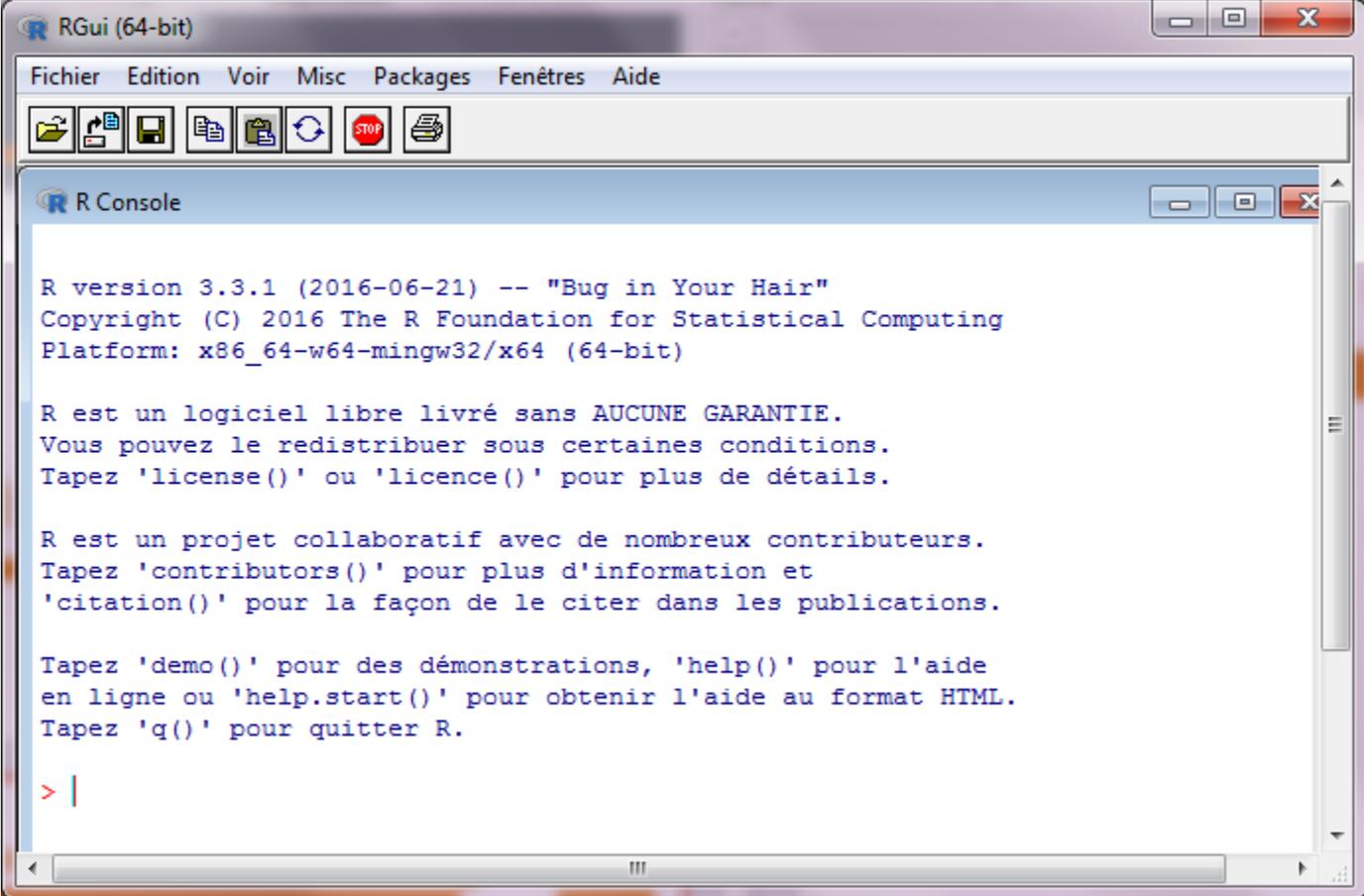
2

- Site de R : <https://www.r-project.org/>
- Manuels : <https://cran.r-project.org/manuals.html>
- Télécharger R <https://www.r-project.org/> ou <https://cran.r-project.org/mirrors.html>

Interface en console

3

- > prompt en attente de commande
- Calculs simples (+-*/)
- Fonctions
- # commentaires
- Language R



```
RGui (64-bit)
Fichier Edition Voir Misc Packages Fenêtres Aide
R Console
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

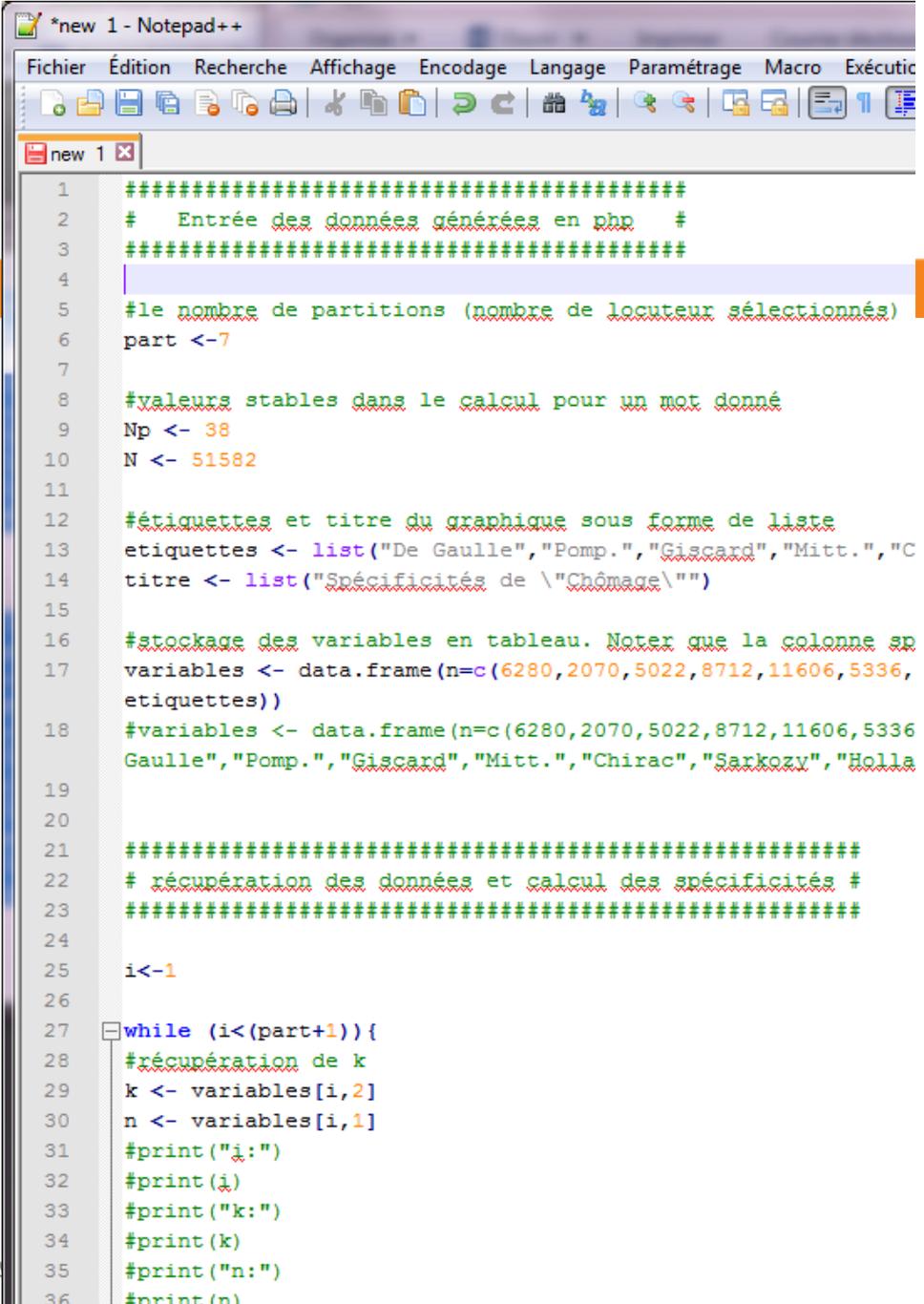
R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> |
```

Notepad++

- Pour visualiser en couleur la syntaxe... (langage > R)
- Insertion dans R par copier-coller
- Permet d'enregistrer des fichier texte qui pourront être lu par R (manipulations redondantes)



The screenshot shows the Notepad++ application window with a menu bar (Fichier, Édition, Recherche, Affichage, Encodage, Langage, Paramétrage, Macro, Exécutic) and a toolbar. The main text area contains R code with syntax highlighting. The code includes comments in French and R commands for data manipulation and printing. The code is as follows:

```
1 #####
2 #  Entrée des données générées en php  #
3 #####
4
5 #le nombre de partitions (nombre de locuteur sélectionnés)
6 part <-7
7
8 #valeurs stables dans le calcul pour un mot donné
9 Np <- 38
10 N <- 51582
11
12 #étiquettes et titre du graphique sous forme de liste
13 etiquettes <- list("De Gaulle", "Pomp.", "Giscard", "Mitt.", "C
14 titre <- list("Spécificités de \"Chômage\")
15
16 #stockage des variables en tableau. Noter que la colonne sp
17 variables <- data.frame(n=c(6280,2070,5022,8712,11606,5336,
18 etiquettes))
19
20 #variables <- data.frame(n=c(6280,2070,5022,8712,11606,5336
21 Gaulle", "Pomp.", "Giscard", "Mitt.", "Chirac", "Sarkozy", "Holla
22 #####
23 # récupération des données et calcul des spécificités #
24 #####
25
26 i<-1
27
28 while (i<(part+1)){
29   #récupération de k
30   k <- variables[i,2]
31   n <- variables[i,1]
32   #print("i:")
33   #print(i)
34   #print("k:")
35   #print(k)
36   #print("n:")
37   #print(n)
```

Construire des graphiques à partir d'une table lexicale



Table lexicale

6

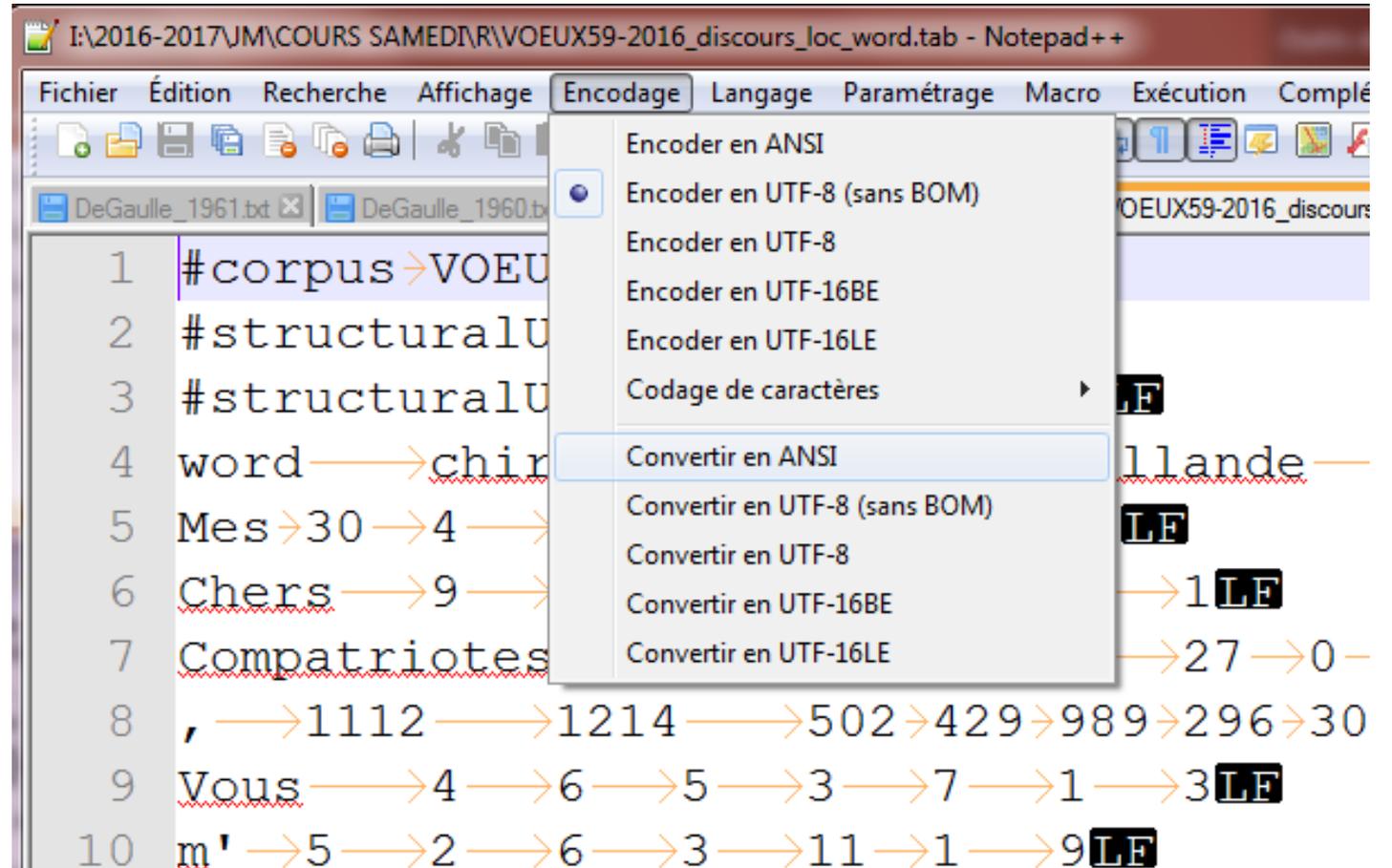
- Exemple d'une table lexicale par locuteur générée par TextObserver

```
IA2016-2017JM\COURS SAMEDI\R\VOEUX59-2016_discours_loc_word.tab - Notepad++
Fichier Édition Recherche Affichage Encodage Langage Paramétrage Macro Exécution Compléments Documents ?
DeGaulle_1961.txt DeGaulle_1960.txt new_1 Voeux_59-2016.xml VOEUX59-2016_discours_loc_word.tab
1 #corpus>VOEUX59-2016CRLF
2 #structuralUnit>discoursCRLF
3 #structuralUnitProperty>locCRLF
4 word->chirac->dg->giscard->hollande->mitterrand->pompidou->sarkozyLF
5 Mes>30->4->3->12->28->0->13LF
6 Chers->9->0->0->2->27->0->1LF
7 Compatriotes->9->0->0->1->27->0->2LF
8 ,->1112->1214->502>429>989>296>306LF
9 Vous->4->6->5->3->7->1->3LF
10 m'>5->2->6->3->11->1->9LF
11 avez->4->5->6->2->5->3->8LF
12 élu>1->0->0->0->1->0->0LF
13 en->147>236>62->95->142>44->68LF
14 mai>1->3->0->1->2->0->1LF
15 dernier>3->5->4->1->4->1->1LF
16 pour->203>123>74->108>122>35->87LF
17 que>155>188>103>74->211>57->104LF
18 nous->181>209>70->78->128>47->79LF
19 constructions->1->0->0->0->0->0->0LF
20 ensemble->19->11->6->2->13->2->2LF
21 une>162>89->54->55->66->19->46LF
22 nouvelle->19->13->4->9->7->1->6LF
23 France->109>83->65->53->82->16->38LF
24 juste->8->0->3->2->3->2->0LF
25 unie->6->3->2->0->0->0->0LF
Normal text file length : 155464 lines : 6699 Ln : 1 Col : 1 Sel : 0 | 0 Dos\Windows UTF-8 w/o BOM INS
```

Encoder en ANSI

7

- Si la table est encodée en UTF-8 (comme ici) choisir :
Convertir en ANSI



Charger la table lexicale

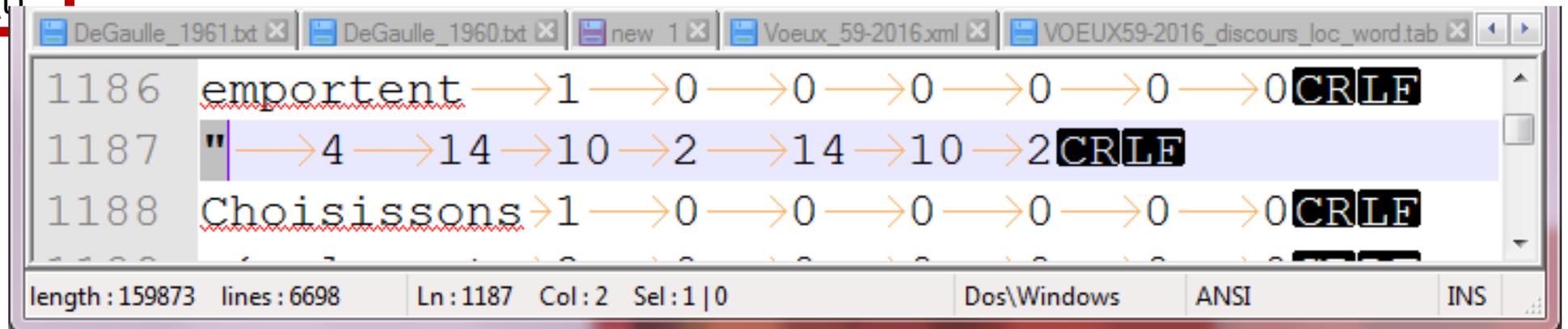
8

- `df <- read.table("C:/Users/marie/Desktop/fichier.tab", sep="\t", header=TRUE, quote = "")`

Indication des séparateurs de colonne (tabulation : \t)

Permet à R de ne pas être gêné par l'entrée " de la table lexicale

Chemin et nom de la table (extension .tab)



The screenshot shows a text editor window with several tabs open. The active tab displays a file with the following content:

```
1186 emportent →1 →0 →0 →0 →0 →0 →0 CR LF
1187 " →4 →14 →10 →2 →14 →10 →2 CR LF
1188 Choisissons →1 →0 →0 →0 →0 →0 →0 CR LF
```

The status bar at the bottom indicates: length : 159873 lines : 6698 Ln : 1187 Col : 2 Sel : 1 | 0. The window title is Dos\Windows and the encoding is ANSI.

Lire l'entête de la table

9

- `>colnames(df)`
- `[1] "word" "chirac" "dg" "giscard" "hollande"`
`"mitterrand" "pompidou" "sarkozy"`

Lire les données

10

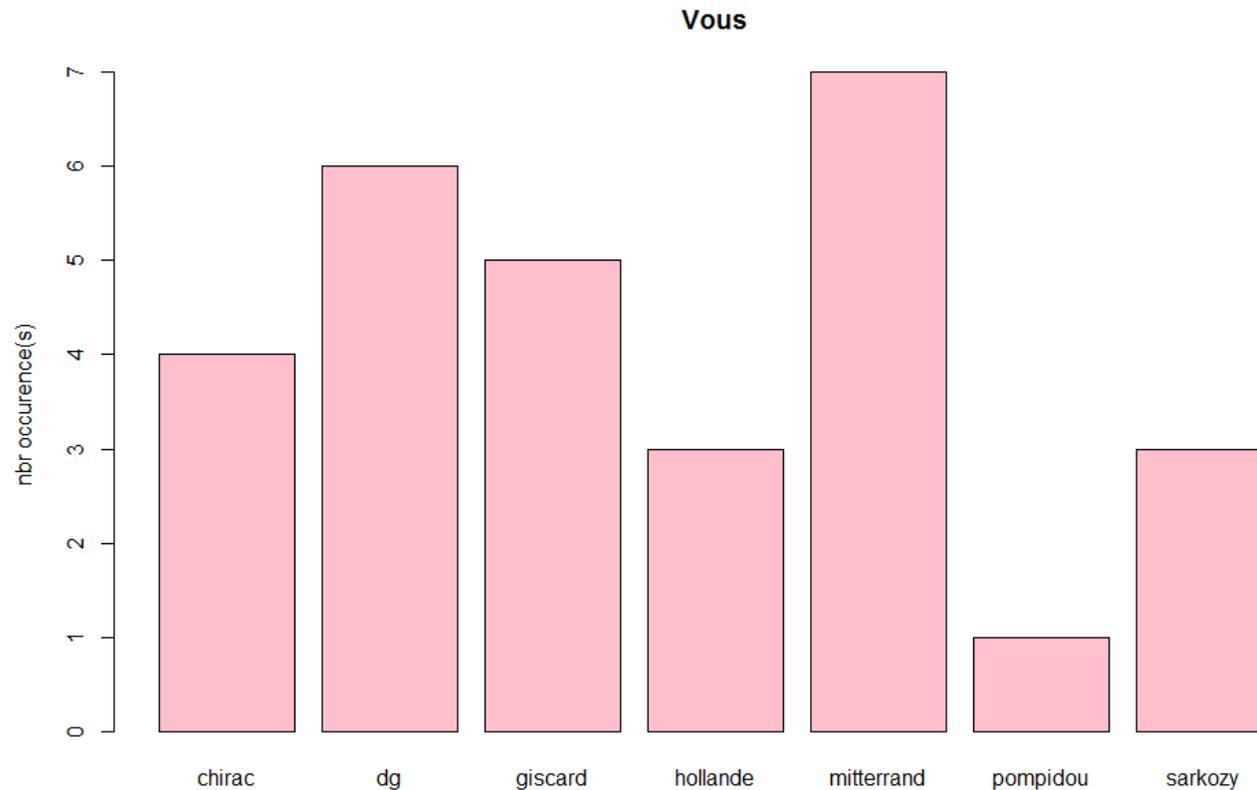
- `> df[5,1]`
- `[1] Vous`
- `> df[5,2]`
- `[1] 4`
- `> colnames(df[2])`
- `[1] "chirac"`

```
DeGaulle_1961.txt DeGaulle_1960.txt new_1 Voeux_59-2016.xml Voeux59-2016_discou
1 #corpus>VOEUX59-2016CRLE
2 #structuralUnit>discoursCRLE
3 #structuralUnitProperty>locCRLE
4 word→chirac→dg→giscard→hollande→
15 Mes→30→4→3→12→28→0→13CRLE
26 Chers→9→0→0→2→27→0→1CRLE
37 Compatriotes→9→0→0→1→27→0→
48 ,→1112→1214→502→429→989→296→30
59 Vous→4→6→5→3→7→1→3CRLE
10 m'→5→2→6→3→11→1→9CRLE
11 avez→4→5→6→2→5→3→8CRLE
12 élu→1→0→0→0→1→0→0CRLE
13 en→147→236→62→95→142→44→68CRLE
```

Répartition de « Vous »

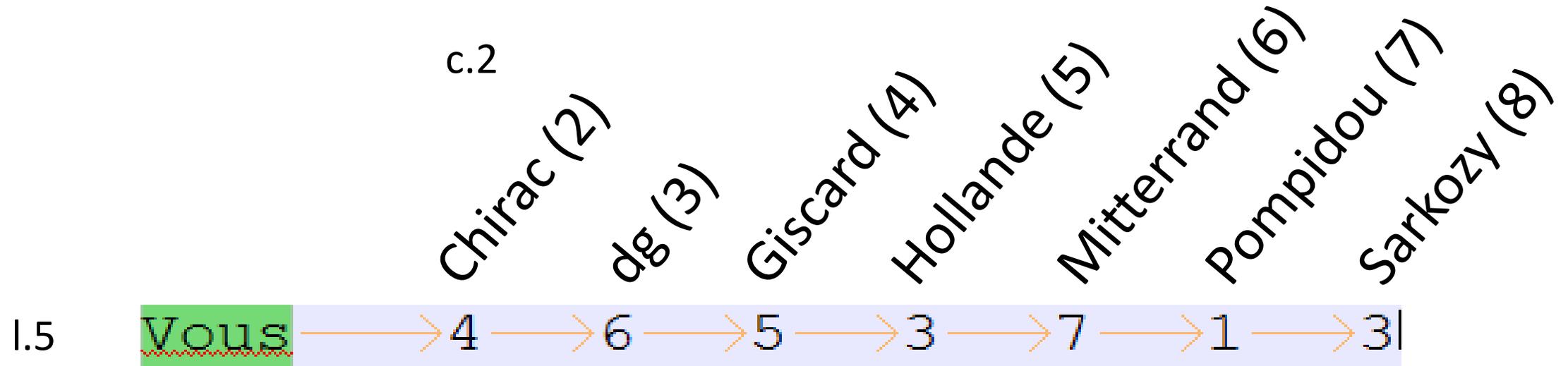
11

- Visualisation du nombre d'occurrence du mot Vous par locuteur



Sélectionner les valeurs

12

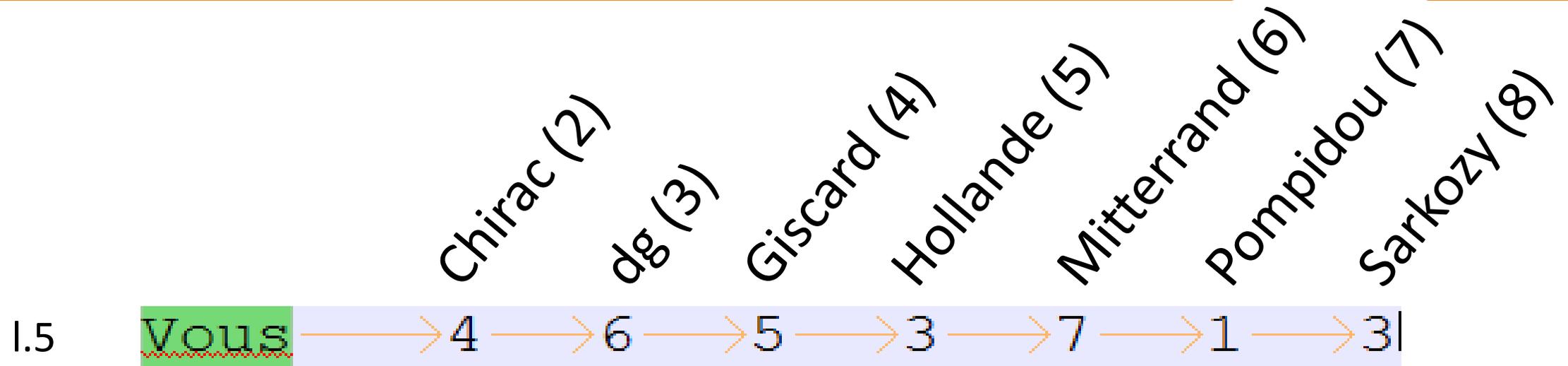


- `x1 <- c(df[5,2],df[5,3],df[5,4],df[5,5],df[5,6],df[5,7],df[5,8])`

Adresse de l'élément : [ligne,colonne]

Sélectionner les étiquettes

13



- A <-
gl(7, 1, 1, labels=c(colnames(df[2]), colnames(df[3]), colnames(df[4]), colnames(df[5]), colnames(df[6]), colnames(df[7]), colnames(df[8])))

Créer un tableau de données



14

- `data <- cbind(x1)`
- `rownames(data) <- levels(A)`

Construire le graphique

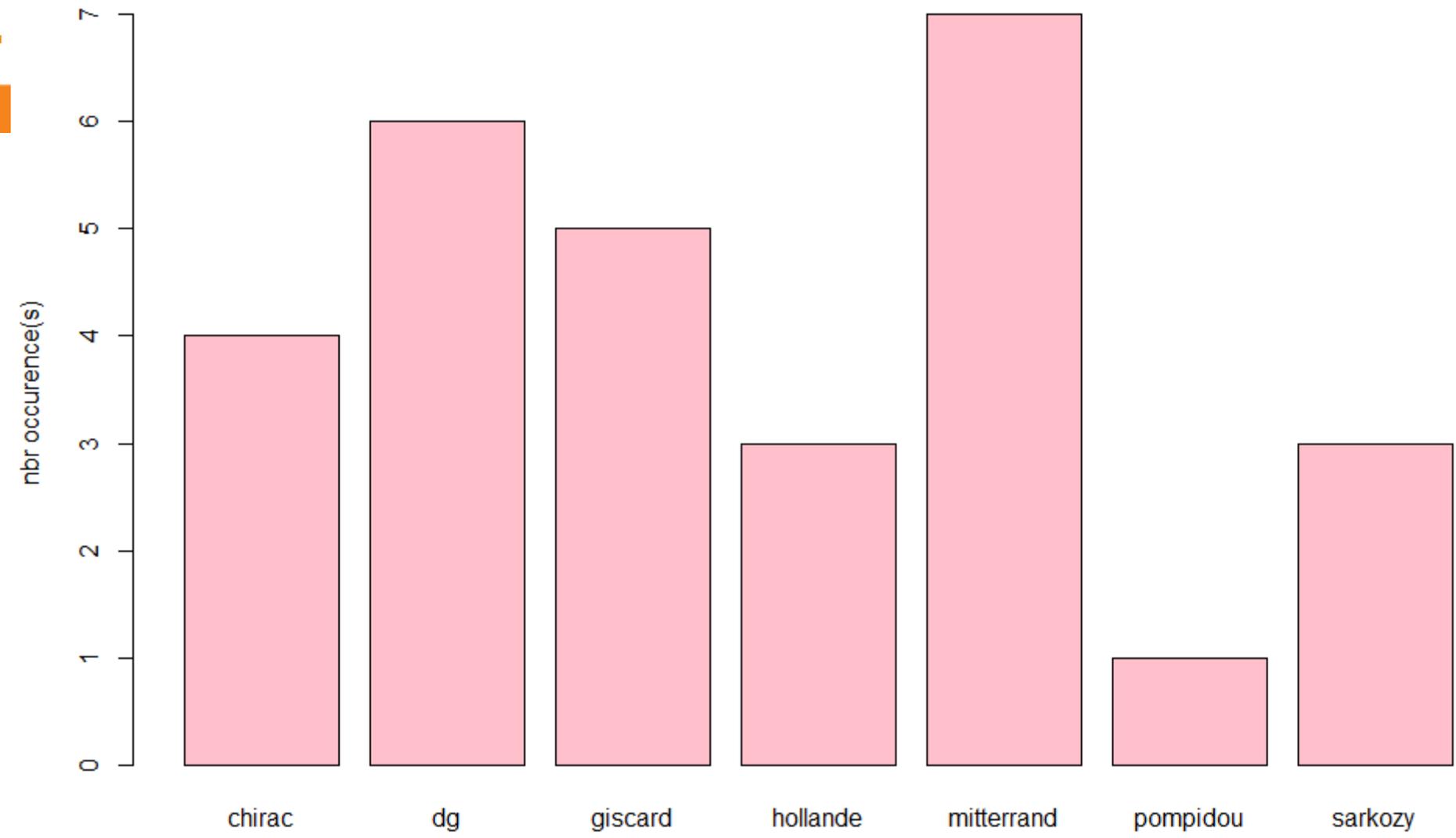
15

- `barplot(x1, names.arg=levels(A), col=c("pink"), ylab="nbr occurrence(s)", main=df[5,1])`



Vous

16



Ensemble du code utilisé

17

- `x1 <-
c(df[5,2],df[5,3],df[5,4],df[5,5],df[5,6],df[5,7],df[5,8])`
- `A <-
gl(7,1,1,labels=c(colnames(df[2]),colnames(df[3]),colnames(df[4]),colnames(df[5]),colnames(df[6]),colnames(df[7]),colnames(df[8])))`
- `data <- cbind(x1)`
- `rownames(data) <- levels(A)`
- `barplot(x1,names.arg=levels(A),col=c("pink"),ylab="nbr
occurrence(s)", main=df[5,1])`

Pour automatiser...



18

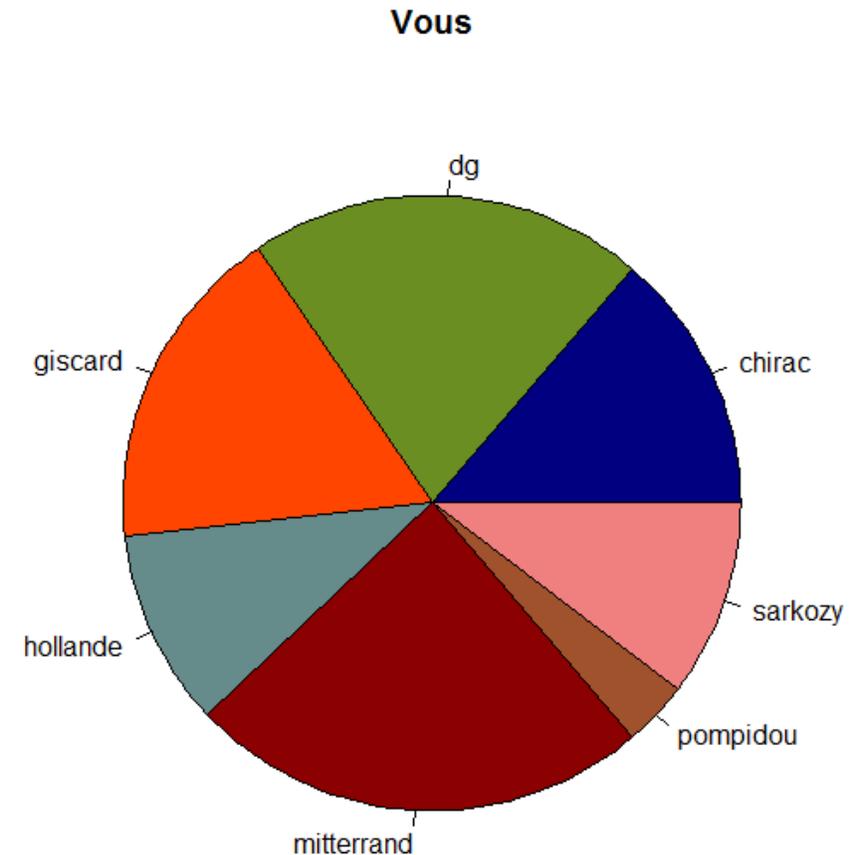
- Boucles
- Conditions

- => on indiquerait à R de chercher le mot « France »
- Mémoriser le mini programme dans un fichier R exécutable

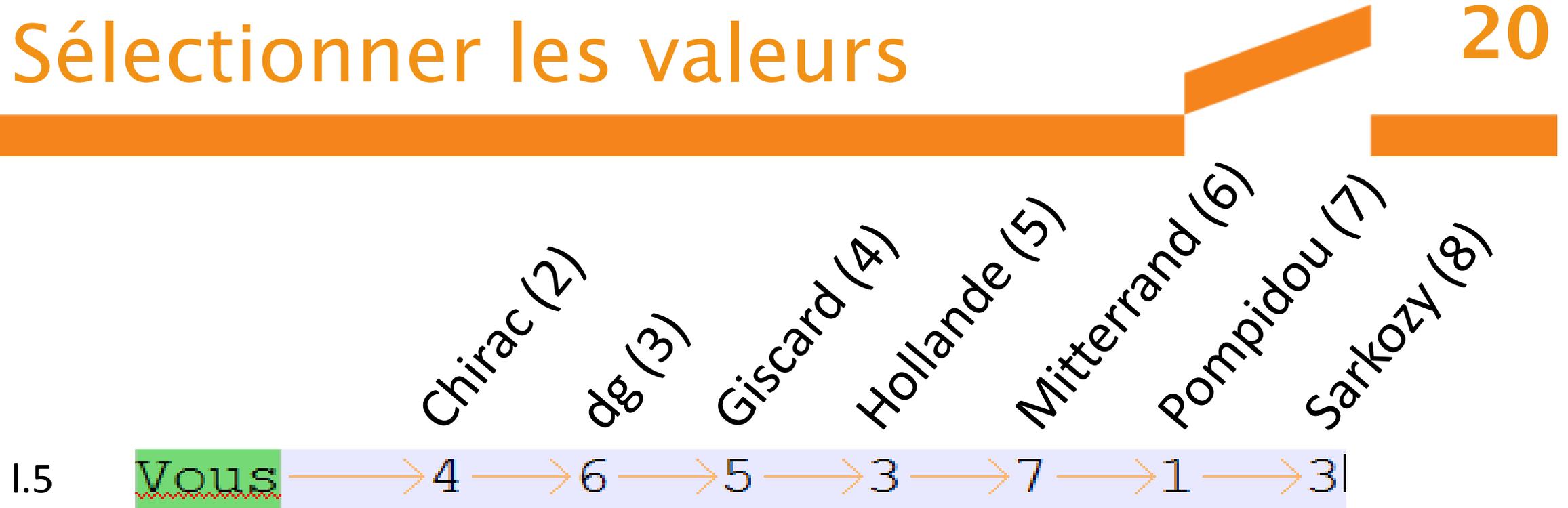
Graphique en secteur

19

- Visualisation du nombre d'occurrence du mot Vous par locuteur



Sélectionner les valeurs



- `data=c(df[5,2],df[5,3],df[5,4],df[5,5],df[5,6],df[5,7],df[5,8])`

Construire le secteur

21

Chirac (2)

dg (3)

Giscard (4)

Hollande (5)

Mitterrand (6)

Pompidou (7)

Sarkozy (8)

1.5

Vous → 4 → 6 → 5 → 3 → 7 → 1 → 3 |

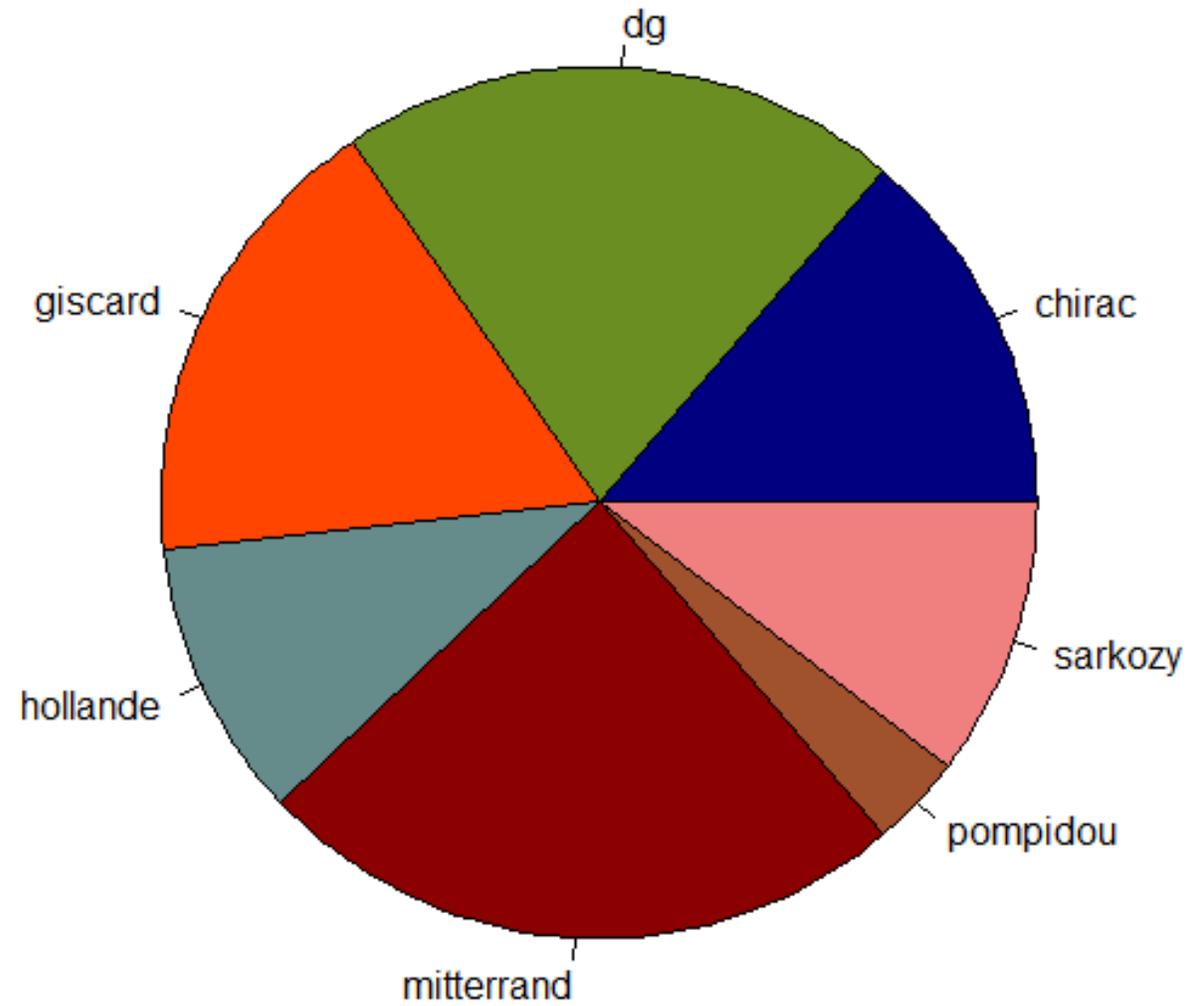
```
pie(data, labels=c(colnames(df[2]),colnames(df[3]),colnames(df[4]),colnames(df[5]),colnames(df[6]),colnames(df[7]),colnames(df[8])), col=c("#000080","#6B8E23","#FF4500","#668B8B","#8B0000","#A0522D","#F08080","#8B2252"), main=df[5,1])
```

Secteur



Vous

2



Exécuter un script R

A decorative graphic consisting of several orange bars. A long horizontal bar spans most of the width of the slide. To its right, there is a shorter horizontal bar. Above the shorter bar, a diagonal bar extends upwards and to the right.

Créer et enregistrer un script

24

- Créer un fichier (txt) avec un blocnote (notepad++ par exemple)
- Encodage ANSI (surtout si on utilise la fonction `print()`)
- L'enregistrer dans l'ordinateur sous le nom : fichier.R (attention l'extension est importante)

Créer le graphique

25

Sur le bureau 2 fichiers : le fichier source des données (voeux2017-loc.tab) et le fichier script R

On lance successivement les deux commandes suivantes :

- `df <- read.table("C:/Users/marie/Desktop/voeux2017-loc.tab", sep="\t", header=TRUE, quote = "")`
//les données sont chargées
- `source(file = "C:/Users/marie/Desktop/graph-barre.R")`
//le script est exécuté à partir des données chargées

Script avec variables

26

- `df <- read.table("C:/Users/marie/Desktop/voeux2017-loc.tab", sep="\t", header=TRUE, quote = "")`
//charge le fichier source (ici le .Tab)
- `source(file = "C:/Users/marie/Desktop/graph-barre.R")`
//lance le script
//reste à taper le numéro de la ligne qui nous intéresse

Enregistrer et modifier des graphiques

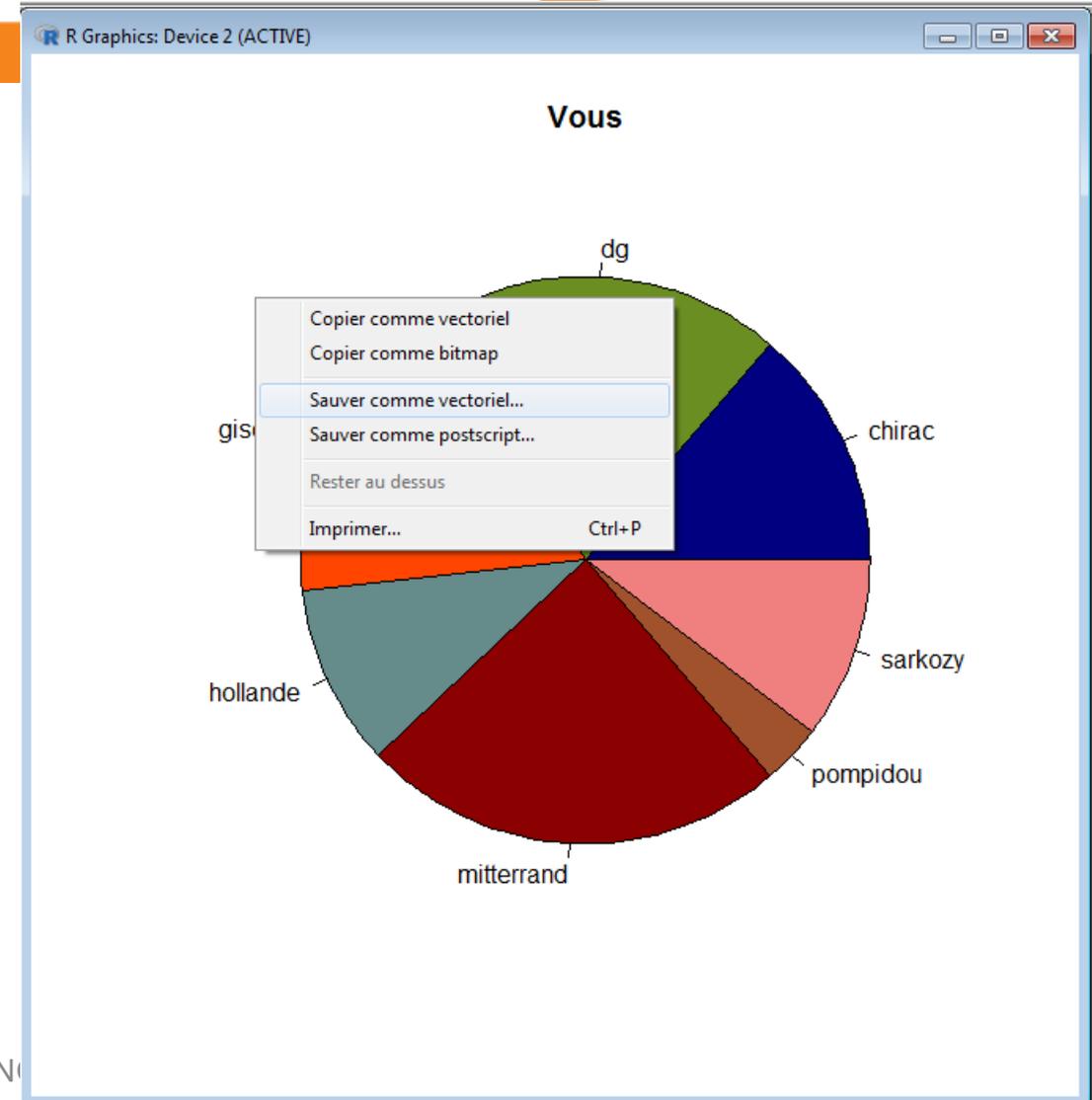


SVG avec Inscape

Enregistrer un graphique

28

- Plusieurs format en fonction de l'usage



Sauver comme vectoriel...
Fichier emf

Inkscape ou Illustrator



29

- Fichier vectoriel : pas de problème de définition
- Modifiable
 - Texte
 - Couleurs
 - Contours
- Exportable en bitmap de haute qualité (édition)

Procédure



30

- Ouvrir le fichier
- Dégroupier
- Sur illustrator : Sélectionner automatiquement les Objets de type Masque et les supprimer...
- Modifier le graphique à sa guise (couleurs, textes, polices)

Références...

31

- <https://cran.r-project.org/>
 - Manuel
- <http://www.statmethods.net/graphs/>
- <https://informatique-mia.inra.fr/r4ciam/node/128>
- https://fr.wikibooks.org/wiki/Programmer_en_R/Programmation_proc%C3%A9durale

