



Project : BigData Processing and Visualization

ZHANG MENG BI2

YAO ZELIANG BI1



1

General description of the data Contentent (pylr)

2

Visualization and insights (ggplot2)

3

R / Shiny

General description of the data content

Import data to R studio

```
> lost<-read.csv2("objets-trouves-gares.csv",encoding = "UTF-8")
> View(lost)
> found<read.csv2("objectfound.csv",encoding = "UTF-8")
```

Basic information of data:(Names of cols ,rows,types...)

```
> attributes(lost)
$names
[1] "Date"           "Type.d.objets"   "Nature.d.objets"
[4] "Type.d.enregistrement" "Gare"           "Code.UIC"

$class
[1] "data.frame"

$row.names
[1] 1 2 3 4 5 6 7 8 9 10 11 12
[16] 16 17 18 19 20 21 22 23 24 25 26 27
[31] 31 32 33 34 35 36 37 38 39 40 41 42
[46] 46 47 48 49 50 51 52 53 54 55 56 57
```

```
> str(lost)
'data.frame': 678775 obs. of 6 variables:
 $ Date : Factor w/ 671150 levels "2013-05-24T1
 $ Type.d.objets : Factor w/ 20 levels "Appareil Ã©lectr
 7 7 ...
 $ Nature.d.objets : Factor w/ 136 levels "Accessoires 2 r
 $ Type.d.enregistrement: Factor w/ 1 level "Déclaration de Per
 $ Gare : Factor w/ 177 levels "", "Abbeville", .
 $ Code.UIC : int 87726000 87726000 87384008 87113
```

General description of the data content

Univariate analysis

```
> summary(lost)
```

Date		Type.d.objets
2014-03-17T18:12:33+01:00:	3	Bagagerie: sacs, valises, cartables :195633
2014-04-18T18:02:11+02:00:	3	Appareils électroniques, informatiques, appareils photo:123332
2014-08-19T16:13:48+02:00:	3	Vêtements, chaussures : 76154
2014-08-25T16:54:29+02:00:	3	Porte-monnaie / portefeuille, argent, titres : 74621
2014-10-22T15:44:11+02:00:	3	Pièces d'identités et papiers personnels : 51943
2014-11-17T17:37:09+01:00:	3	Optique : 36483
(other)	:678757	(Other) :120609

Nature.d.objets	Type.d.enregistrement	Gare
Porte-monnaie, portefeuille	: 64997	Déclaration de Perte:678775 :450352
Téléphone portable	: 60719	Paris Gare de Lyon: 20844
Sac à dos	: 45999	Paris Montparnasse: 20068
Valise, sac sur roulettes	: 41606	Paris Gare du Nord: 16880
Manteau, veste, blazer, parka, blouson, cape	: 37977	Paris Saint-Lazare: 13656
Sac de voyage, sac de sport, sac à bandoulière:	36181	Strasbourg : 13218
(other)	:391296	(other) :143757

Code.UIC
Min. :87111849
1st Qu.:87271007
Median :87444877
Mean :87470127
3rd Qu.:87686006
Max. :87785006
NA's :450352

General description of the data content

Extract some data & filter:

```
> library(plyr)
> attach(llost)
> Total_TypeObject_Lost<-count(Type.d.objets)
```

TypeObject_Lost	Total
3 Appareils électroniques, informatiques, appareils photo	123332
4 Articles d'enfants, de puériculture	10002
5 Articles de sport, loisirs, camping	6075
6 Articles médicaux	3816
7 Bagagerie: sacs, valises, cartables	195633
8 Bijoux, montres	7936
9 Clés, porte-clés, badge magnétique	29819
10 Divers	17501
11 Instruments de musique	1877
12 Keys, keyholders, magnetic badges	1
13 Livres, articles de papeterie	28695
14 Optique	36483
15 Parapluies	5662
16 Pièces d'identités et papiers personnels	51943
17 Porte-monnaie / portefeuille, argent, titres	74621
18 Vélos, trottinettes, accessoires 2 roues	4396
19 Vélos, trottinettes, accessoires 2 roues	4827
20 Vêtements, chaussures	76154



General description of the data content

```
> Total_Lost_Found<-merge(Count_Lost,Count_Found,all.y= TRUE)
```

106	Mo<U+00FB>tiers - Salins - Brides-les-Bains	10	60
16	Argenteuil	3	55
22	Bar-le-Duc	2	7
97	lyon perrache	2	6
178	Bergerac	NA	8
179	Forbach	NA	1
180	Lille Flandres	NA	9
181	Marne-la-Vallée	NA	10

Problems found



Visualization and insights (ggplot2)



Visualization and insights (ggplot2)

Select 100 random samples (start with small amount of data)

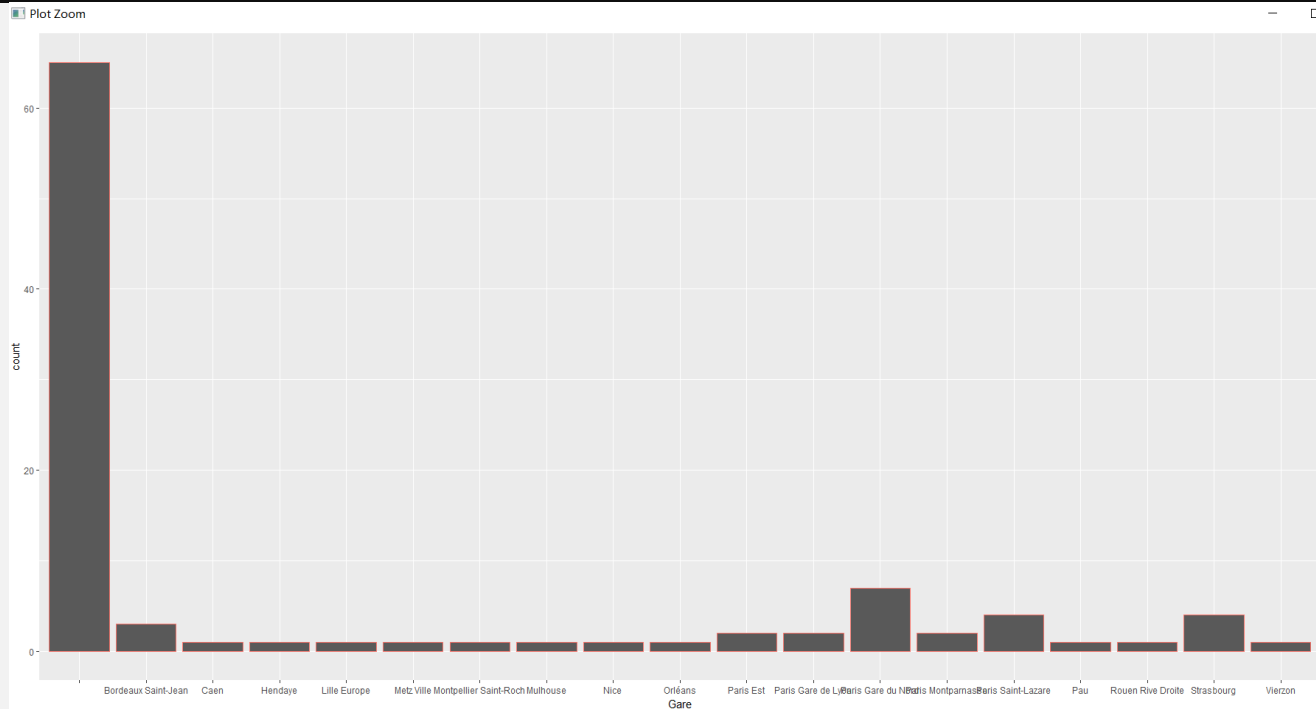
```
> lost_small = lost[sample(nrow(lost), 100), ]  
> found_small = found[sample(nrow(found), 100), ]  
>
```

	Date	Type.d.objets	Nature.d.objets
354273	2017-08-30T13:12:37+02:00	Vêtements, chaussures	Pull over, gilet
223208	2013-06-14T20:14:35+02:00	Porte-monnaie / portefeuille, argent, titres	Porte-monnaie, portefeuille
613065	2015-03-23T10:28:00+01:00	Bagagerie: sacs, valises, cartables	Attaché-case, serviette
380297	2016-11-15T19:41:50+01:00	Bagagerie: sacs, valises, cartables	Autre bagagerie (préciser)
33981	2016-04-03T23:00:32+02:00	Vêtements, chaussures	Autres vêtements (préciser)
381192	2015-11-02T18:26:58+01:00	Bagagerie: sacs, valises, cartables	Sacoche pour ordinateur, photo, vidéo
655521	2017-11-11T09:08:49+01:00	Porte-monnaie / portefeuille, argent, titres	Porte-monnaie, portefeuille
488015	2014-08-22T12:58:12+02:00	Bagagerie: sacs, valises, cartables	Nécessaire, mallette de toilette, sac de maqu
273918	2016-12-05T23:20:52+01:00	Bagagerie: sacs, valises, cartables	Autre bagagerie (préciser)
251188	2014-08-25T17:54:40+02:00	Vêtements, chaussures	Manteau, veste, blazer, parka, blouson, cape
426729	2017-01-17T19:18:49+01:00	Vêtements, chaussures	Manteau, veste, blazer, parka, blouson, cape
445664	2017-03-03T11:48:59+01:00	Vêtements, chaussures	Foulard, écharpe



Visualization and insights (ggplot2) *Histogram with ggplot2*

```
ggplot(Gare, data=lost_small, color="red")
```



```
ggplot(lost_small, aes(x=Gare, y=Type.d.objets))+geom_point()+  
geom_smooth()+geom_abline()
```

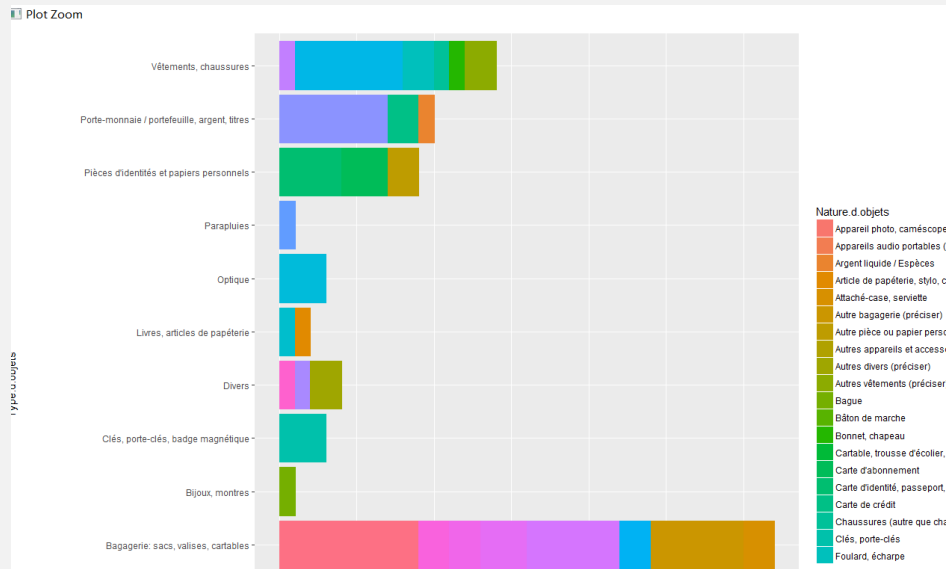




Visualization and insights (ggplot2)

Barplot with ggplot2 in lost data_set

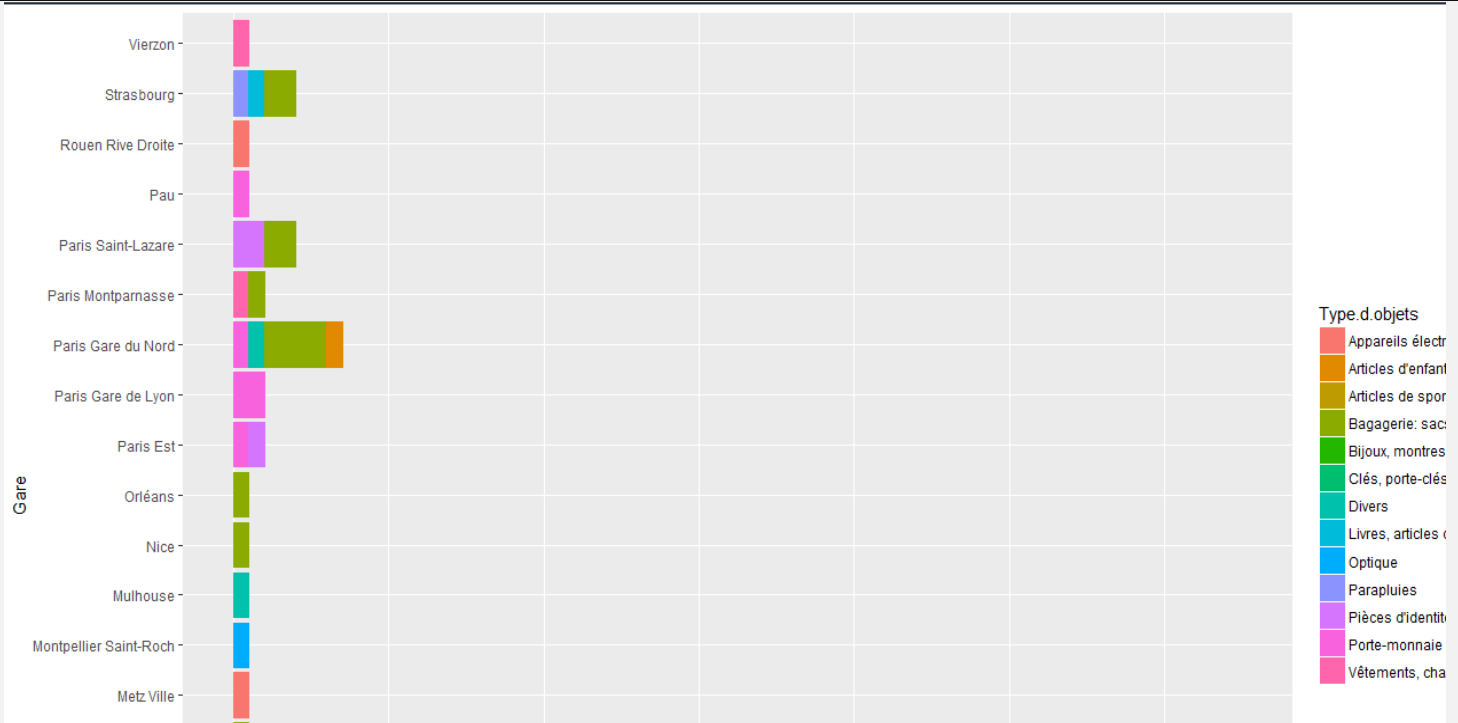
```
ggplot(Type.d.objets, data=lost_small, geom="bar", colour=Nature.d.objets, fill=Nature.d.objets)+coord_flip()
```



Visualization and insights (ggplot2)

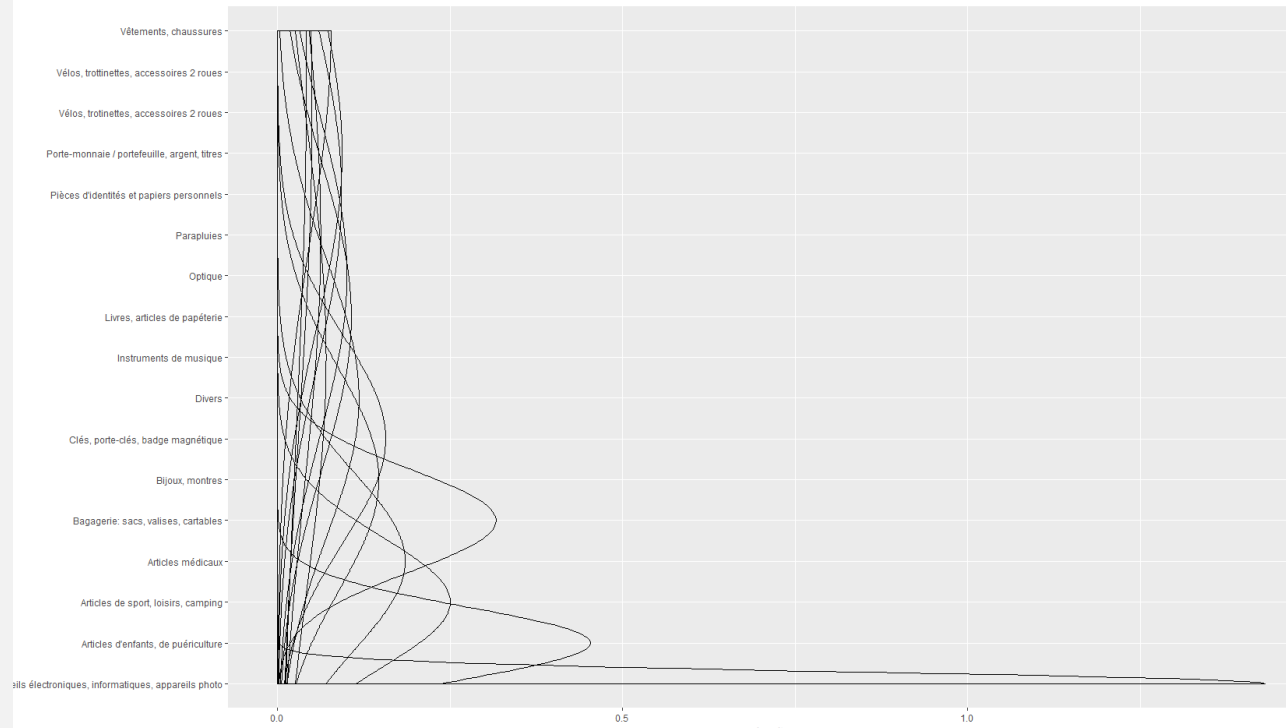
Barplot with ggplot2 in lost data_set

```
ggplot(Gare, data=lost_small, geom="bar", colour=Type.d.objets, fill=Type.d.objets)
+coord_flip()
```



Visualization and insights (ggplot2) *Densitygram with ggplot2*

```
> qplot(Type.d.objets,data=lost_Medium,geom="density")+coord_flip()
```





R / Shiny