homework ii

Mick McQuaid

2018-01-01

Introduction

(Introductory text should go here.)

Initialization

Here we load the tidyverse packages and the `data.table` package and load the nyc311 data set. Then we fix the column names of the nyc311 data so that they have no spaces.

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------- tidyverse 1.2.1 --
```

```
## ✓ ggplot2 3.0.0     ✓ purrr   0.2.5
## ✓ tibble  1.4.2     ✓ dplyr   0.7.6
## ✓ tidyr   0.8.1     ✓ stringr 1.3.1
## ✓ readr   1.1.1     ✓ forcats 0.3.0
```

```
## -- Conflicts -------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(data.table)
```

```
## data.table 1.11.6  Latest news: r-datatable.com
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
nyc311<-fread("311_Service_Requests_from_2010_to_Present.csv")
names(nyc311)<-names(nyc311) %>%
  stringr::str_replace_all("\\s", ".")
```

Description

Here we describe the data, showing both a sample and a data dictionary.

The head of the table

Here we produce a table of just some relevant columns of data.

```
library(xtable)
options(xtable.comment=FALSE)
options(xtable.booktabs=TRUE)
narrow<-nyc311 %>%
  select(Agency,
     Complaint.Type,
     Descriptor,
     Incident.Zip,
     Status,
     Borough)
xtable(head(narrow))
```

	Agency	Complaint.Type	Descriptor	Incident.Zip	Status	Borough
1	NYPD	Vending	In Prohibited Area	10465	Closed	BRONX
2	NYPD	Blocked Driveway	No Access	11234	Open	BROOKLYN
3	NYPD	Noise - Street/Sidewalk	Loud Music/Party	11204	Open	BROOKLYN
4	NYPD	Noise - Street/Sidewalk	Loud Talking	11211	Assigned	BROOKLYN
5	NYPD	Noise - Street/Sidewalk	Loud Talking	10025	Closed	MANHATTAN
6	NYPD	Noise - Street/Sidewalk	Loud Talking	11205	Closed	BROOKLYN

Data Dictionary

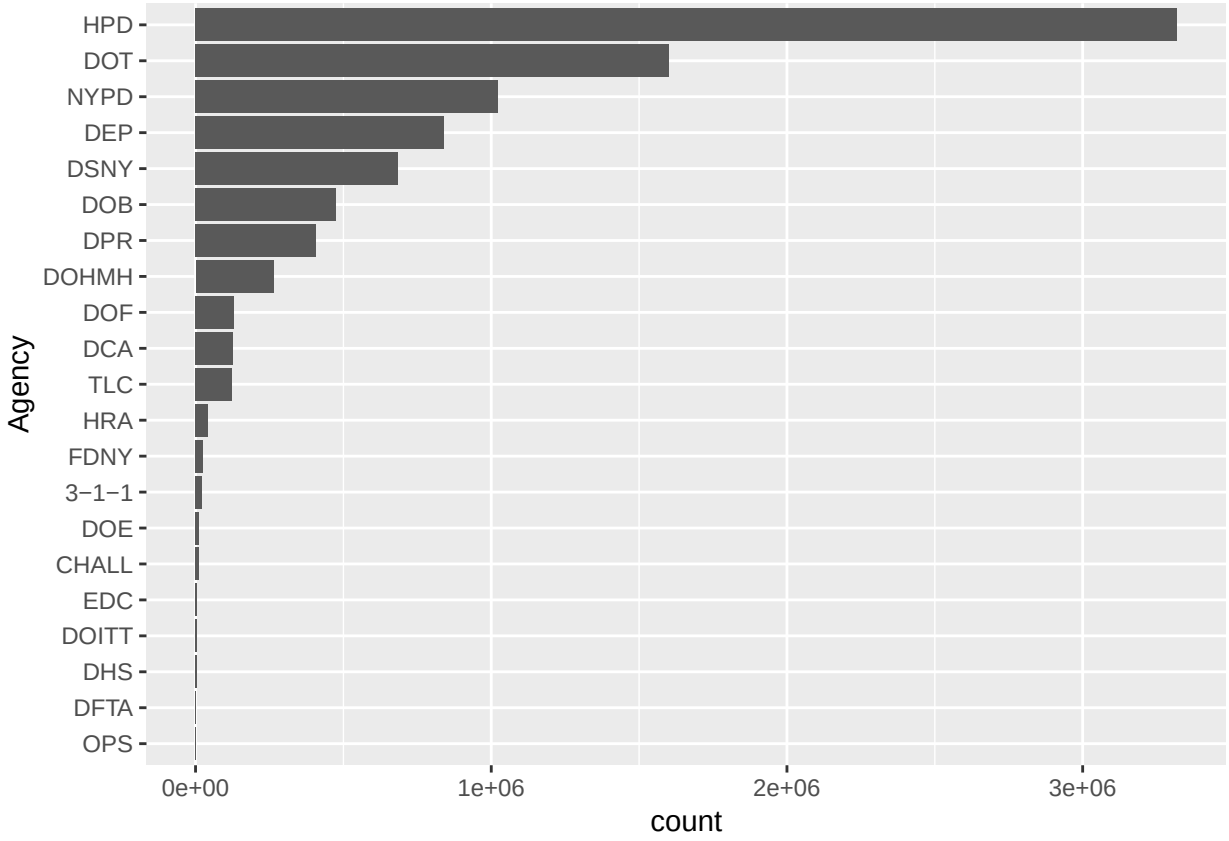
(Here is where you should put the data dictionary.)

Exploration

Here we explore the columns in the data set.

(A description of the following plot should go here.)

```
bigAgency <- narrow %>%
  group_by(Agency) %>%
  summarize(count=n()) %>%
  filter(count>1000)
bigAgency$Agency<-factor(bigAgency$Agency,
  levels=bigAgency$Agency[order(bigAgency$count)])
p<-ggplot(bigAgency,aes(x=Agency,y=count)) +
   geom_bar(stat="identity") +
   coord_flip()
p
```



(More plots should follow here.)

Next we include a crosstabulation.

```
xtabA<-dplyr::filter(narrow,
  Complaint.Type=='HEATING' |
  Complaint.Type=='GENERAL CONSTRUCTION' |
  Complaint.Type=='PLUMBING'
)
xtabB<-select(xtabA,Borough,"Complaint.Type")
library(gmodels)
CrossTable(xtabB$Borough,xtabB$'Complaint.Type')
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1868064
##
##
##                | xtabB$Complaint.Type
## xtabB$Borough | GENERAL CONSTRUCTION |              HEATING |            PLUMBING |            Row Total |
## --------------|----------------------|----------------------|----------------------|----------------------|
##         BRONX |               107626 |               195246 |               103964 |               406836 |
##               |               23.326 |               19.145 |                1.030 |                      |
##               |                0.265 |                0.480 |                0.256 |                0.218 |
##               |                0.215 |                0.220 |                0.217 |                      |
```

```
##               |                0.058 |                0.105 |                0.056 |                      |
## --------------|----------------------|----------------------|----------------------|----------------------|
##      BROOKLYN |               132552 |               190268 |               128383 |               451203 |
##               |             1076.405 |             2717.190 |             1398.387 |                      |
##               |                0.294 |                0.422 |                0.285 |                0.242 |
##               |                0.264 |                0.214 |                0.268 |                      |
##               |                0.071 |                0.102 |                0.069 |                      |
## --------------|----------------------|----------------------|----------------------|----------------------|
##     MANHATTAN |                61453 |               137458 |                63103 |               262014 |
##               |             1123.330 |             1347.582 |              245.877 |                      |
##               |                0.235 |                0.525 |                0.241 |                0.140 |
##               |                0.123 |                0.155 |                0.132 |                      |
##               |                0.033 |                0.074 |                0.034 |                      |
## --------------|----------------------|----------------------|----------------------|----------------------|
##        QUEENS |                41277 |                75776 |                43604 |               160657 |
##               |               79.707 |                4.192 |              142.183 |                      |
##               |                0.257 |                0.472 |                0.271 |                0.086 |
##               |                0.082 |                0.085 |                0.091 |                      |
##               |                0.022 |                0.041 |                0.023 |                      |
## --------------|----------------------|----------------------|----------------------|----------------------|
## STATEN ISLAND |                 8329 |                6011 |                 7525 |                21865 |
##               |             1030.062 |             1845.525 |              657.654 |                      |
##               |                0.381 |                0.275 |                0.344 |                0.012 |
##               |                0.017 |                0.007 |                0.016 |                      |
##               |                0.004 |                0.003 |                0.004 |                      |
## --------------|----------------------|----------------------|----------------------|----------------------|
##   Unspecified |               150277 |               282916 |               132296 |               565489 |
##               |               15.587 |              750.862 |             1106.709 |                      |
##               |                0.266 |                0.500 |                0.234 |                0.303 |
##               |                0.300 |                0.319 |                0.276 |                      |
##               |                0.080 |                0.151 |                0.071 |                      |
## --------------|----------------------|----------------------|----------------------|----------------------|
##  Column Total |               501514 |               887675 |               478875 |              1868064 |
##               |                0.268 |                0.475 |                0.256 |                      |
## --------------|----------------------|----------------------|----------------------|----------------------|
##
##
```

(Some discussion of the above crosstab should follow.)

(More crosstabs or corrplots should follow.)

Conclusion

(Tell what you did in this document here.)