myAnalysis

Shaodong Wang

2017 年 2 月 5 日

First, import the data(like what professor did).

```
library(plyr)
library(choroplethr)
```

```
## Warning: package 'choroplethr' was built under R version 3.3.2
```

```
## Loading required package: acs
```

```
## Warning: package 'acs' was built under R version 3.3.2
```

```
## Loading required package: stringr
```

```
## Loading required package: XML
```

```
## Warning: package 'XML' was built under R version 3.3.2
```

```
##
## Attaching package: 'acs'
```

```
## The following object is masked from 'package:base':
##
##     apply
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.3.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:acs':
##
##     combine
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

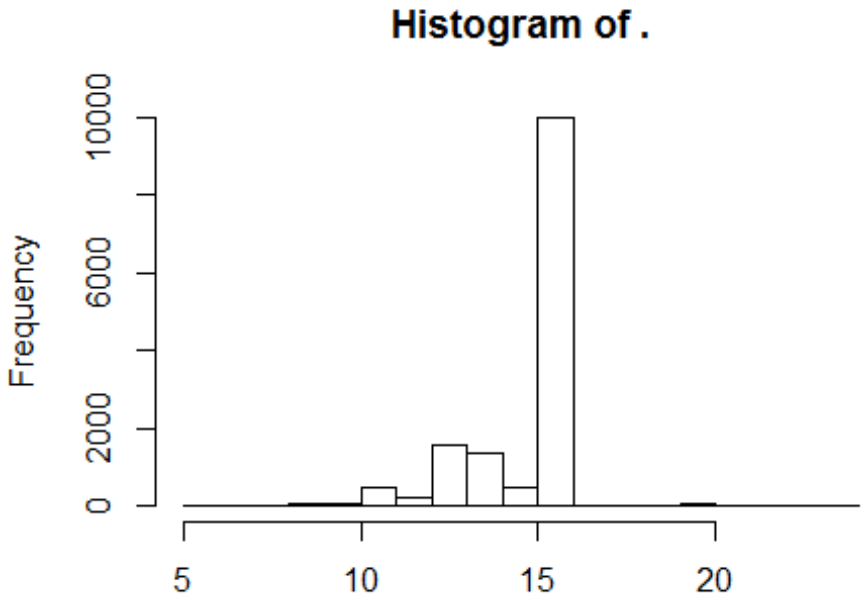
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(readr)

## Warning: package 'readr' was built under R version 3.3.2

library(data.table)

## Warning: package 'data.table' was built under R version 3.3.2

## ------------------------------------------------------------------------
-----

## data.table + dplyr code now lives in dtplyr.
## Please library(dtplyr)!

## ------------------------------------------------------------------------
-----

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

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

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.3.2

## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: purrr

## Warning: package 'ggplot2' was built under R version 3.3.2

## Warning: package 'tibble' was built under R version 3.3.2

## Warning: package 'tidyr' was built under R version 3.3.2

## Warning: package 'purrr' was built under R version 3.3.2

## Conflicts with tidy packages ----------------------------------------
-------

## arrange():   dplyr, plyr
## between():   dplyr, data.table
## combine():   dplyr, acs
## compact():   purrr, plyr
## count():     dplyr, plyr
## failwith():  dplyr, plyr
```

```
## filter():    dplyr, stats
## first():     dplyr, data.table
## id():        dplyr, plyr
## lag():       dplyr, stats
## last():      dplyr, data.table
## mutate():    dplyr, plyr
## rename():    dplyr, plyr
## summarise(): dplyr, plyr
## summarize(): dplyr, plyr
## transpose(): purrr, data.table

dest = "https://www.fhwa.dot.gov/bridge/nbi/2016/delimited/WI16.txt"
WI16 = fread(dest)

## Warning in fread(dest): Bumped column 125 to type character on data
row
## 11570, field contains '00PE093'. Coercing previously read values in
this
## column from logical, integer or numeric back to character which may
not
## be lossless; e.g., if '00' and '000' occurred before they will now b
e just
## '0', and there may be inconsistencies with treatment of ',,' and ',N
A,' too
## (if they occurred in this column before the bump). If this matters p
lease
## rerun and set 'colClasses' to 'character' for this column. Please no
te that
## column type detection uses a sample of 1,000 rows (100 rows at 10 po
ints)
## so hopefully this message should be very rare. If reporting to datat
able-
## help, please rerun and include the output from verbose=TRUE.

WI16 = as.tbl(WI16)
classes = sapply(WI16, class)
```

Then, exclude the strange data(missing value, all 0, etc)

```
M = WI16
M = M[,-14]
is.na(M) %>% rowSums %>% hist
```

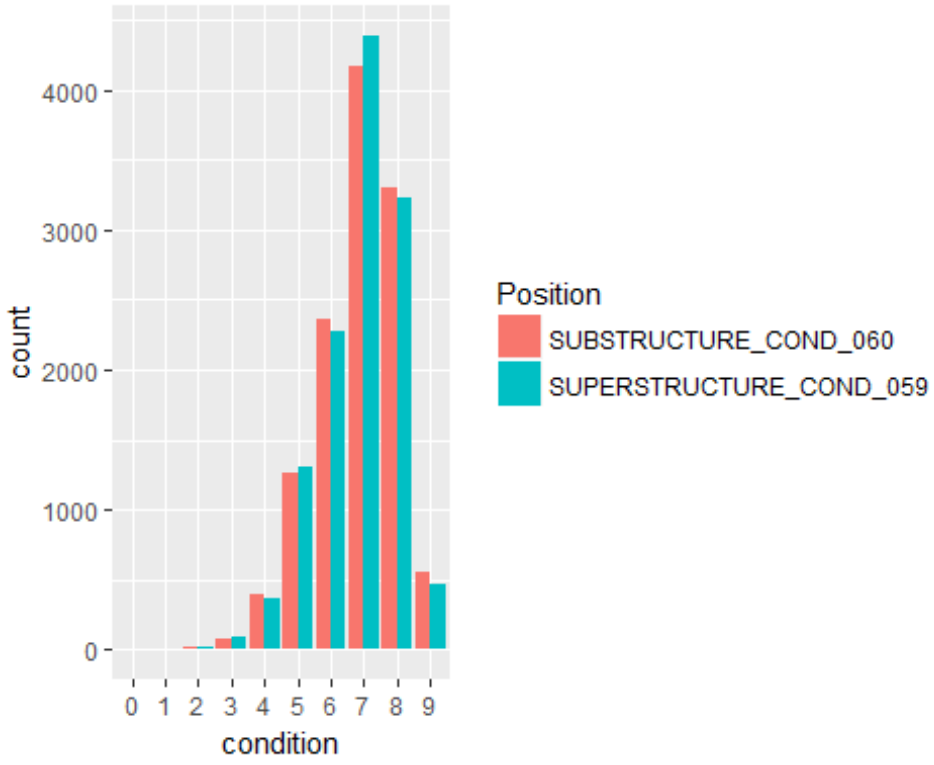



```
fun = function(x){ return(which(x>20)) }
(bad =  is.na(M) %>% colSums %>% fun)

##  CRITICAL_FACILITY_006B        SUBROUTE_NO_013B    OPERATING_RATING_
064
##                      12                      18
 72
##    INVENTORY_RATING_066      WORK_PROPOSED_075A       WORK_DONE_BY_0
75B
##                      74                      81
 82
## FRACTURE_LAST_DATE_093A UNDWATER_LAST_DATE_093B     SPEC_LAST_DATE_0
93C
##                      89                      90
 91
##     BRIDGE_IMP_COST_094    ROADWAY_IMP_COST_095      TOTAL_IMP_COST_
096
##                      92                      93
 94
##         YEAR_OF_IMP_097   OTHER_STATE_CODE_098A   OTHER_STATE_PCNT_0
98B
##                      95                      96
 97
##      TEMP_STRUCTURE_103   PERCENT_ADT_TRUCK_109     PIER_PROTECTION_
111
##                     102                     110
112
```

```
##      MIN_NAV_CLR_MT_116                  REMARKS              PROJ_SUF
FIX
##                     117                      122
125
##         NBI_TYPE_OF_IMP         DTL_TYPE_OF_IMP            SPECIAL_C
ODE
##                     126                      127
128
##               STEP_CODE
##                     129

M = M[,-bad]
M = as.tbl(M)
```

Select variables that I am interested in.

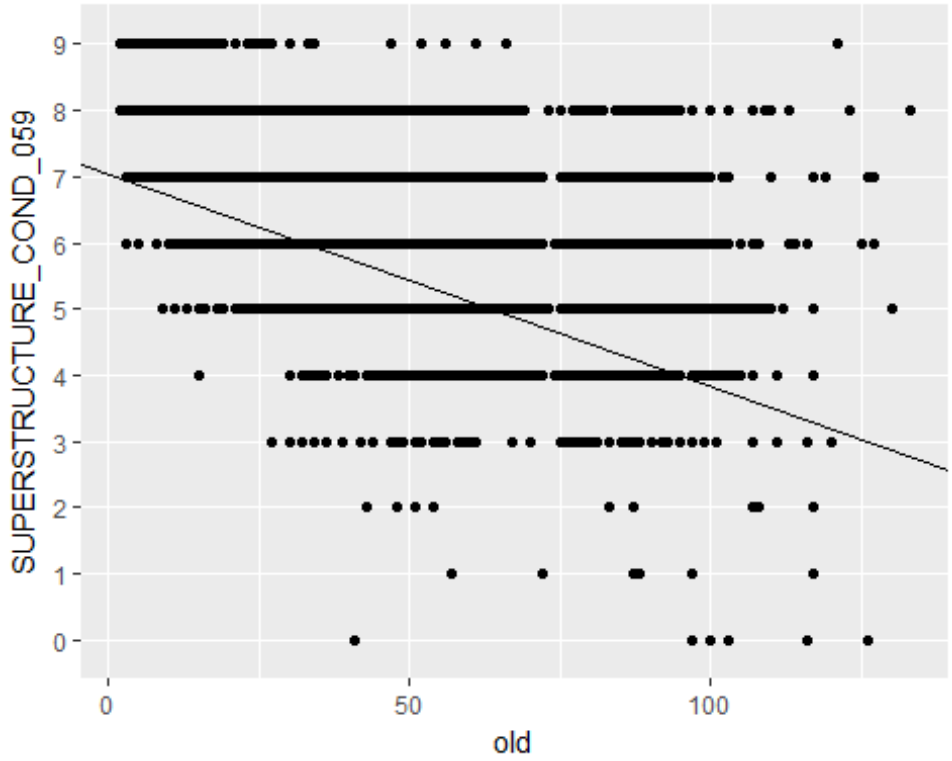
```
keep=c("STATE_CODE_001", "STRUCTURE_NUMBER_008" ,"YEAR_BUILT_027","SUPE
RSTRUCTURE_COND_059", "SUBSTRUCTURE_COND_060" )
my = select(M, one_of(keep))
mywi = filter(my,SUPERSTRUCTURE_COND_059 !="N",SUBSTRUCTURE_COND_060 !=
"N")
mywi=mutate(mywi, old = 2017- YEAR_BUILT_027,super_Sub_compare=as.numer
ic(SUPERSTRUCTURE_COND_059)-as.numeric(SUBSTRUCTURE_COND_060))
mywi2<-mywi%>%gather(SUPERSTRUCTURE_COND_059,SUBSTRUCTURE_COND_060,key=
Position,value=condition)
ggplot(data=mywi2)+geom_bar(mapping=aes(x=condition,fill=Position),posi
tion = "dodge")
```



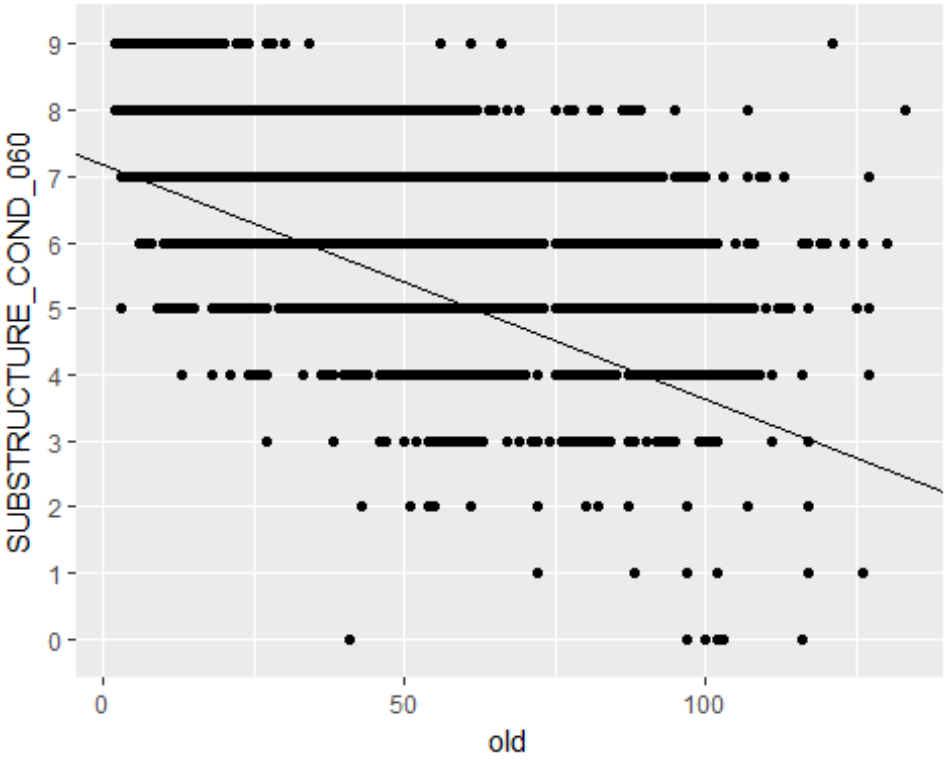
```
table(mywi$super_Sub_compare)

##
##   -6   -4   -3   -2   -1    0    1    2    3    4    5
##    1    9   97  532 2304 6424 2138  523   91   11    2
```

This result shows that superstructure is not more durable than substructure, vice versa.

```
library(ggplot2)
m1=lm(SUPERSTRUCTURE_COND_059~old,data=mywi)
m2=lm(SUBSTRUCTURE_COND_060~old,data=mywi)
ggplot(data=mywi)+geom_point(mapping = aes(x=old,y=SUPERSTRUCTURE_COND_
059))+geom_abline(aes(intercept=m1$coefficients[1],slope=m1$coefficient
s[2]))
```




```
ggplot(data=mywi)+geom_point(mapping = aes(x=old,y=SUBSTRUCTURE_COND_06
0))+geom_abline(aes(intercept=m2$coefficients[1],slope=m2$coefficients
[2]))
```



These two plots both shows the older the bridge is, the worse the condition is(just like our common sense).But there're some bridge reconstructed after their built_year, so some bridges older than 100 years are still in good condition.