Playground

Frances Hung

10/21/2017

Motivation

According to the CDC, suicide was the 10th leading cause of death in the US in 2015, and the 2nd leading cause of death among adolescents and young adults. Psychological disorders, particularly depression, are a significant risk factor for suicide especially when they go untreated. There is no reliable way to predict who is at risk for committing suicide, because most screening approaches depend on self-report information and people contemplating on suicide would often deny it when asked. However, even if someone wouldn't tell the truth on a questionnaire, they will often tell Google. Using suicide rate and mental health treatment facilities data as well as Google search term data, our project aims to map the demand for and supply of mental health treatment in California cities.

- use result/visualization as hook

Playground

```
require(gtrendsR)
```

```
## Loading required package: gtrendsR
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(dplyr)
```

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

```
require(zipcode)
```

```
## Loading required package: zipcode
```

```
data("zipcode")
require(ggmap)
```

```
## Loading required package: ggmap
```

Making Dataframes

This gives us a master dataframe of search frequencies of "depression" over the past 12 months in the US which relate for sure to mental health. We can take different dataframes using "$": see the dataframe for details.

```
trend<-gtrends("suicide",c("US"),time="today+5-y")
trend$interest_by_region
```

```
##                location hits keyword geo gprop
## 1               Wyoming  100 suicide  US   web
## 2                Alaska  100 suicide  US   web
## 3            New Mexico   99 suicide  US   web
## 4                Nevada   97 suicide  US   web
## 5                  Utah   96 suicide  US   web
## 6               Montana   95 suicide  US   web
## 7               Arizona   93 suicide  US   web
## 8               Indiana   93 suicide  US   web
## 9         West Virginia   92 suicide  US   web
## 10                Idaho   91 suicide  US   web
## 11              Vermont   91 suicide  US   web
## 12             Colorado   90 suicide  US   web
## 13                Maine   89 suicide  US   web
## 14         South Dakota   89 suicide  US   web
## 15           California   88 suicide  US   web
## 16             Delaware   88 suicide  US   web
## 17        New Hampshire   88 suicide  US   web
## 18             Nebraska   88 suicide  US   web
## 19             Kentucky   88 suicide  US   web
## 20             Arkansas   87 suicide  US   web
## 21             Missouri   87 suicide  US   web
## 22             Oklahoma   86 suicide  US   web
## 23         Pennsylvania   86 suicide  US   web
## 24           Washington   86 suicide  US   web
## 25             Michigan   86 suicide  US   web
## 26                 Iowa   85 suicide  US   web
## 27         North Dakota   85 suicide  US   web
## 28                 Ohio   85 suicide  US   web
## 29                Texas   84 suicide  US   web
## 30         Rhode Island   83 suicide  US   web
## 31             Illinois   83 suicide  US   web
## 32           New Jersey   83 suicide  US   web
## 33          Connecticut   82 suicide  US   web
## 34 District of Columbia   82 suicide  US   web
## 35            Tennessee   82 suicide  US   web
## 36              Alabama   81 suicide  US   web
## 37        Massachusetts   81 suicide  US   web
## 38                Hawaii   81 suicide  US   web
## 39             Maryland   81 suicide  US   web
## 40               Kansas   80 suicide  US   web
## 41            Wisconsin   80 suicide  US   web
## 42            Louisiana   79 suicide  US   web
## 43       South Carolina   78 suicide  US   web
## 44       North Carolina   78 suicide  US   web
## 45            Minnesota   77 suicide  US   web
```

```
## 46          Mississippi   76 suicide  US   web
## 47             New York   76 suicide  US   web
## 48              Georgia   76 suicide  US   web
## 49              Florida   75 suicide  US   web
## 50             Virginia   70 suicide  US   web
## 51               Oregon   70 suicide  US   web
```

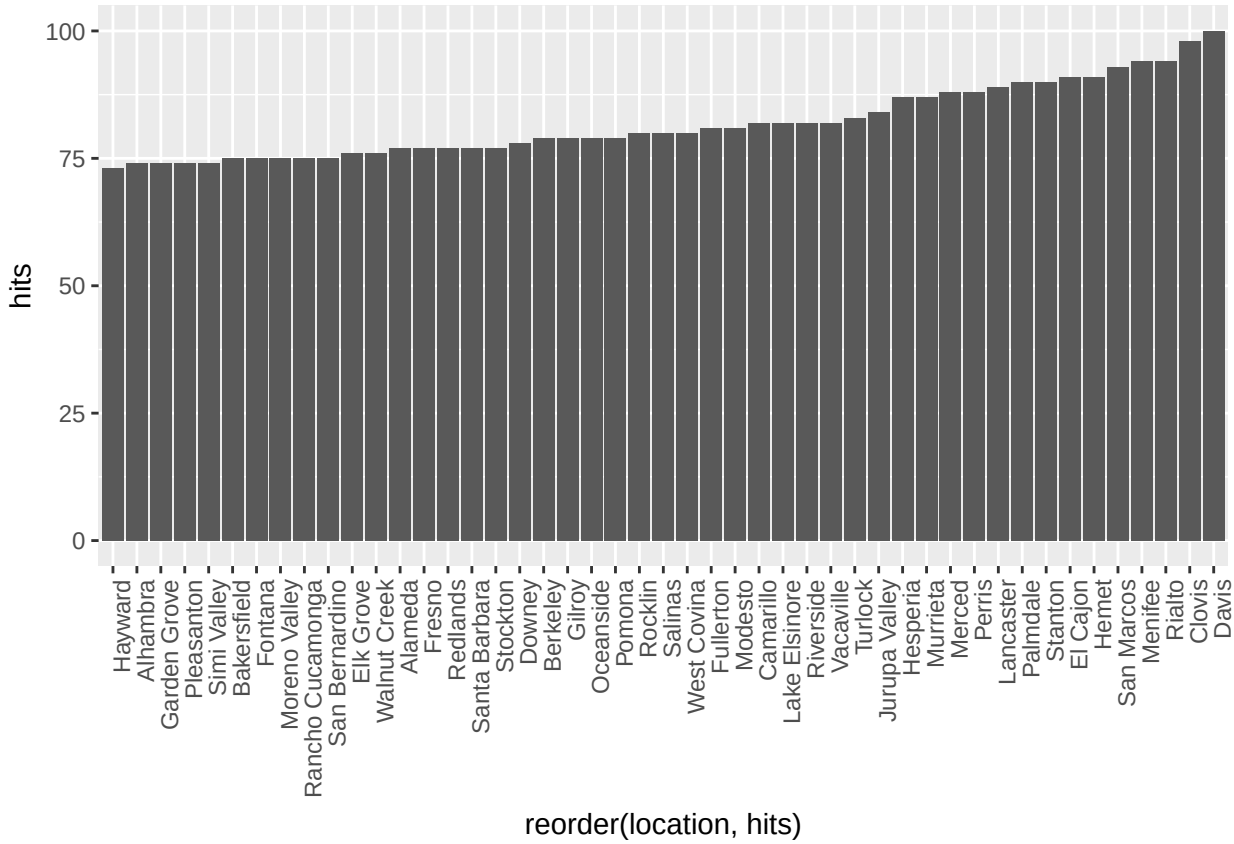
For example, this gives us search frequencies by cities in CA in the U.S.

```
cities_longlat<-read.csv("cal_cities.csv",header=TRUE) %>% select(c(location,Latitude,Longitude))
cities_dep<-gtrends("depression",c("US-CA"),time="today 12-m")$interest_by_city
cities_dep<-cities_dep %>% inner_join(cities_longlat,by="location")

write.csv(cities_dep,file="cities_top49.csv")
```

This plots cities_dep.

```
ggplot(cities_dep,aes(x=reorder(location,hits),y=hits))+geom_bar(stat="identity")+theme(axis.text.x = el
```




```
# for (i in 1:length(cities_dep$location)) {
#   place=geocode(cities_dep$location[i],output="latlon",source="dsk")
#   cities_dep$lat[i]=as.numeric(place[1])
#   cities_dep$lon[i]=as.numeric(place[2])
#   print(place)
# }
cities_dep$keyword<-NULL
cities_dep$geo<-NULL
cities_dep$gprop<-NULL
```

```
# center=as.numeric(geocode("United States",source="dsk"))
# mappy<-get_map(c(-119.4179,36.7783),zoom=6,scale=2,maptype = "terrain",source="google")
# p=ggmap(mappy,extent="device",ylab="Latitude",xlab="Longitude")
# p=p+geom_point(data=cities_dep,aes(x=lat,y=lon),size=(cities_dep$hits/50)^2)
# p
```

```
suis<-read.csv(file="death.csv",header=TRUE) %>% filter(Causes.of.Death=="SUI") #%>% filter(Year >= 201
colnames(suis)[2]<-"zip"
suis$zip<-as.character(suis$zip)
city_suis<-inner_join(zipcode,suis,by="zip")
city_suis<-city_suis %>% group_by(city) %>% summarise(suicides=sum(Count))
colnames(city_suis)[1]<-"location"
```

```
citydem<-read.csv("citydems.csv",header=TRUE)
citydem2<-read.csv("citydems2.csv",header=TRUE)
citydem2$Name<-gsub(",.*","",citydem2$Name)
citydem$Name<-gsub(",.*","",citydem$Name)
citydem$FIPS<-NULL
citydem2$FIPS<-NULL
colnames(citydem)<-c("location", "male", "female","healthcare","bluecollar","whitecollar","nonfamily","
colnames(citydem2)<-c("location","healthcarepp","activities","socialRec","entertainment","pov","presdrug
logtable<-inner_join(citydem,cities_dep,by="location") %>% inner_join(city_suis,by="location") %>% inne
#logtable$hiRate<- ifelse(logtable$suicides>median(logtable$suicides),1,0)
logtable_crop<-logtable[,-c(1,23,24)]
```

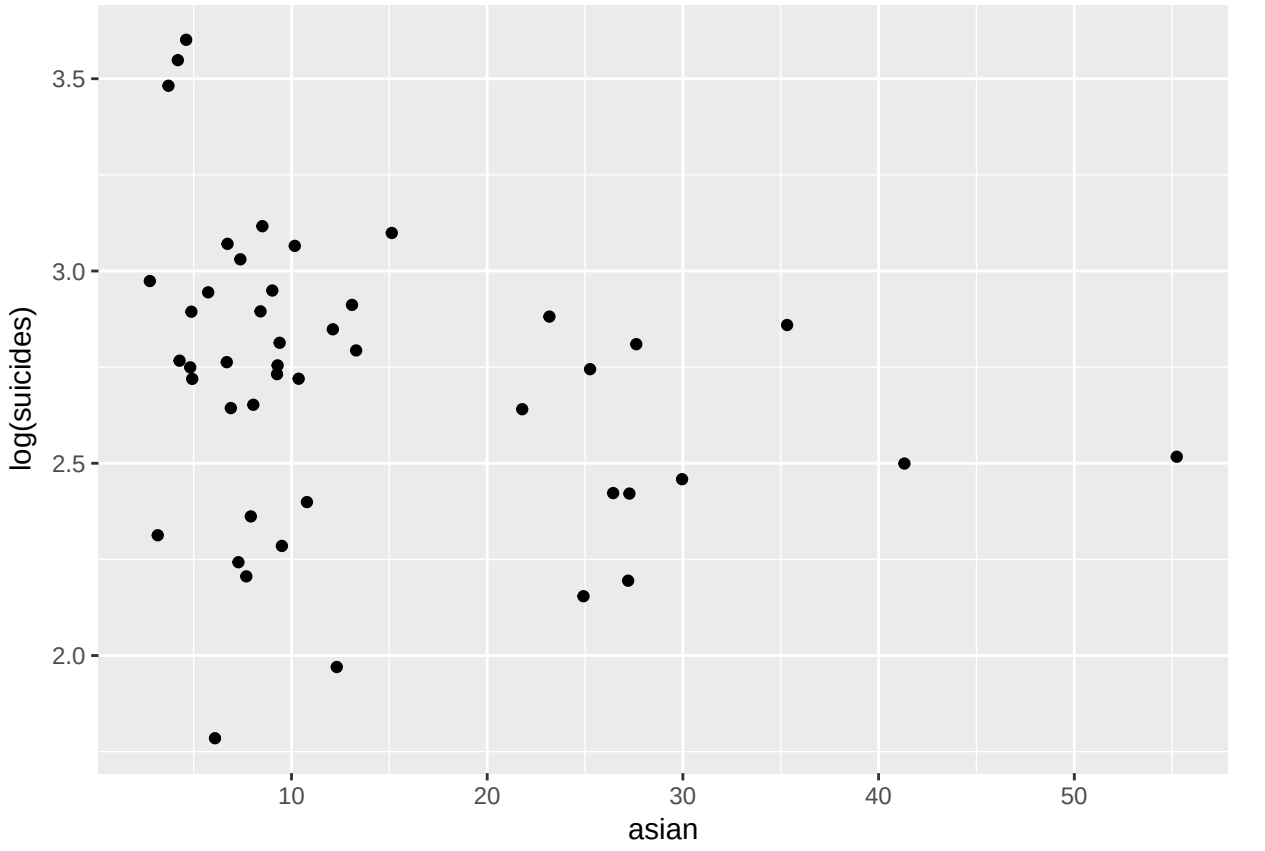
```
set.seed(1)
model<-lm(log(suicides)~.,data=logtable_crop)
summary(model)
```

```
##
## Call:
## lm(formula = log(suicides) ~ ., data = logtable_crop)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54014 -0.09323  0.00481  0.11368  0.36579
##
## Coefficients: (1 not defined because of singularities)
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.104e+00  1.095e+01  -0.557   0.5841
## male           1.697e-02  8.747e-02   0.194   0.8483
## female                NA         NA      NA       NA
## healthcare     3.986e-05  1.423e-03   0.028   0.9780
## bluecollar     2.630e-02  2.508e-02   1.049   0.3081
## whitecollar   -2.337e-02  3.150e-02  -0.742   0.4676
## nonfamily      3.834e-02  4.930e-02   0.778   0.4469
## medAge         1.602e-01  1.233e-01   1.299   0.2104
## AmInd          1.126e-01  2.044e-01   0.551   0.5886
## whiteNonHisp  -5.812e-04  6.883e-02  -0.008   0.9934
## hisp          -1.799e-02  5.818e-02  -0.309   0.7607
## white         -4.144e-02  3.591e-02  -1.154   0.2636
## black         -7.980e-02  7.461e-02  -1.070   0.2990
## asian         -5.853e-02  5.768e-02  -1.015   0.3236
## medIncome      4.589e-05  3.653e-05   1.256   0.2251
```

```
## lessHS         2.293e-03  3.749e-02   0.061   0.9519
## HS             2.505e-04  4.230e-02   0.006   0.9953
## Bachelors      2.065e-02  3.645e-02   0.566   0.5781
## pop            2.179e-08  8.188e-07   0.027   0.9791
## unmarriedMpop -1.966e-02  4.619e-02  -0.426   0.6754
## unemployed    -4.342e-02  7.237e-02  -0.600   0.5560
## hits           7.748e-03  1.227e-02   0.632   0.5355
## healthcarepp  -6.590e+00  2.779e+01  -0.237   0.8152
## activities    -2.423e+02  5.396e+02  -0.449   0.6588
## socialRec     -2.703e+02  1.641e+02  -1.647   0.1169
## entertainment  4.210e+01  2.767e+01   1.521   0.1455
## pov            8.875e-02  3.145e-02   2.822   0.0113 *
## presdrugs     -1.003e+02  1.900e+02  -0.528   0.6039
## healthcarebiz  2.448e-01  1.331e-01   1.839   0.0825 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2858 on 18 degrees of freedom
## Multiple R-squared:  0.7737, Adjusted R-squared:  0.4342
## F-statistic: 2.279 on 27 and 18 DF,  p-value: 0.03661
```

Variables significant in this regression are poverty rate (+), density of healthcare businesses(+), and black/Asian population percentwise (-) in the city. Variables which also should be considered according to this regression are percentage of white-collar workers (-), searches of "depression" (+), median income (+), hispanice population (-), and white non-Hispanic population (-).

```
ggplot(logtable_crop,aes(x=asian,y=log(suicides)))+geom_point()
```



```r
library(caret)
```

```
## Loading required package: lattice
```

```r
modelrf<-train(suicides~.,method="rf",tuneGrid=data.frame(mtry=c(2,3,4,5,6)),data=logtable_crop)
```

```
## randomForest 4.6-12
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
## The following object is masked from 'package:dplyr':
##
##     combine
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
modelrf$finalModel
```

```
##
## Call:
##  randomForest(x = x, y = y, mtry = param$mtry)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 6
##
##           Mean of squared residuals: 31.01858
##                     % Var explained: 23.78
```

facilities data

```r
# facilities <- read.csv("filtered_licensed-healthcare-facility-listing-june-30-2017.csv")
# facilities <- filter(facilities, LICENSE_CATEGORY_DESC == "Acute Psychiatric Hospital"|LICENSE_CATEGO
# View(facilities)
# write.csv(facilities, file="facilities.csv")
# # more facilities
# facilities.2 <- read_csv("facilities.csv")
# View(facilities.2)
# write.csv(facilities.2, file = "facilities_2.csv")
```