Google reviews

Slučajni šumari

02 05 2020

Učitavanje podataka i deskriptivna statistika

```
df <- read_csv("data.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   User = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
df[df == 0] <- NA
head(df)
```

```
## # A tibble: 6 x 25
##   User   churches resorts beaches parks theatres museums malls  zoo
##   <chr>     <dbl>   <dbl>   <dbl> <dbl>    <dbl>   <dbl> <dbl> <dbl>
## 1 User~        NA      NA    3.63  3.65        5    2.92     5  2.35
## 2 User~        NA      NA    3.63  3.65        5    2.92     5  2.64
## 3 User~        NA      NA    3.63  3.63        5    2.92     5  2.64
## 4 User~        NA     0.5    3.63  3.63        5    2.92     5  2.35
## 5 User~        NA      NA    3.63  3.63        5    2.92     5  2.64
## 6 User~        NA      NA    3.63  3.63        5    2.92     5  2.63
## # ... with 16 more variables: restaurants <dbl>, `pubs/bars` <dbl>, `local
## #   services` <dbl>, `burger/pizza shops` <dbl>, `hotels/other
## #   lodgings` <dbl>, `juice bars` <dbl>, `art galleries` <dbl>, `dance
## #   clubs` <dbl>, `swimming pools` <dbl>, gyms <dbl>, bakeries <dbl>,
## #   `beauty & spas` <dbl>, cafes <dbl>, `view points` <dbl>,
## #   monuments <dbl>, gardens <dbl>
```

```
glimpse(df)
```

```
## Observations: 5,456
## Variables: 25
## $ User                 <chr> "User 1", "User 2", "User 3", "User 4"...
## $ churches             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ resorts              <dbl> NA, NA, NA, 0.50, NA, NA, 5.00, 5.00, ...
## $ beaches              <dbl> 3.63, 3.63, 3.63, 3.63, 3.63, 3.63, 3....
## $ parks                <dbl> 3.65, 3.65, 3.63, 3.63, 3.63, 3.63, 3....
## $ theatres             <dbl> 5.00, 5.00, 5.00, 5.00, 5.00, 5.00, 5....
## $ museums              <dbl> 2.92, 2.92, 2.92, 2.92, 2.92, 2.92, 2....
## $ malls                <dbl> 5.00, 5.00, 5.00, 5.00, 5.00, 5.00, 3....
## $ zoo                  <dbl> 2.35, 2.64, 2.64, 2.35, 2.64, 2.63, 2....
## $ restaurants          <dbl> 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2....
## $ `pubs/bars`          <dbl> 2.64, 2.65, 2.64, 2.64, 2.64, 2.65, 2....
## $ `local services`     <dbl> 1.70, 1.70, 1.70, 1.73, 1.70, 1.71, 1....
## $ `burger/pizza shops` <dbl> 1.69, 1.69, 1.69, 1.69, 1.69, 1.69, 1....
```

```
## $ `hotels/other lodgings` <dbl> 1.70, 1.70, 1.70, 1.70, 1.70, 1.69, 1....
## $ `juice bars`            <dbl> 1.72, 1.72, 1.72, 1.72, 1.72, 1.72, 1....
## $ `art galleries`         <dbl> 1.74, 1.74, 1.74, 1.74, 1.74, 1.74, 1....
## $ `dance clubs`           <dbl> 0.59, 0.59, 0.59, 0.59, 0.59, 0.59, 0....
## $ `swimming pools`        <dbl> 0.50, 0.50, 0.50, 0.50, 0.50, 0.50, 0....
## $ gyms                    <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ bakeries                <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5...
## $ `beauty & spas`         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ cafes                   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ `view points`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ monuments               <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ gardens                 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
```

`summary(df)`

```
##      User              churches         resorts           beaches
##  Length:5456        Min.   :0.50    Min.   :0.500    Min.   :0.500
##  Class :character   1st Qu.:0.97    1st Qu.:1.370    1st Qu.:1.540
##  Mode  :character   Median :1.35    Median :1.970    Median :2.065
##                     Mean   :1.51    Mean   :2.359    Mean   :2.491
##                     3rd Qu.:1.86    3rd Qu.:2.720    3rd Qu.:2.740
##                     Max.   :5.00    Max.   :5.000    Max.   :5.000
##                     NA's   :195     NA's   :90       NA's   :4
##      parks           theatres         museums           malls
##  Min.   :0.830    Min.   :1.120    Min.   :1.110    Min.    :1.120
##  1st Qu.:1.730    1st Qu.:1.770    1st Qu.:1.790    1st Qu.:1.930
##  Median :2.460    Median :2.670    Median :2.680    Median :3.230
##  Mean   :2.797    Mean   :2.959    Mean   :2.893    Mean   :3.351
##  3rd Qu.:4.093    3rd Qu.:4.312    3rd Qu.:3.840    3rd Qu.:5.000
##  Max.   :5.000    Max.   :5.000    Max.   :5.000    Max.    :5.000
##
##      zoo          restaurants       pubs/bars        local services
##  Min.   :0.860    Min.   :0.840    Min.   :0.810    Min.   :0.78
##  1st Qu.:1.620    1st Qu.:1.800    1st Qu.:1.640    1st Qu.:1.58
##  Median :2.170    Median :2.800    Median :2.680    Median :2.00
##  Mean   :2.541    Mean   :3.126    Mean   :2.833    Mean   :2.55
##  3rd Qu.:3.190    3rd Qu.:5.000    3rd Qu.:3.530    3rd Qu.:3.22
##  Max.   :5.000    Max.   :5.000    Max.   :5.000    Max.   :5.00
##                                                     NA's   :1
##  burger/pizza shops hotels/other lodgings   juice bars      art galleries
##  Min.   :0.780      Min.   :0.770         Min.   :0.760    Min.   :0.500
##  1st Qu.:1.290      1st Qu.:1.190         1st Qu.:1.030    1st Qu.:0.860
##  Median :1.690      Median :1.610         Median :1.490    Median :1.330
##  Mean   :2.078      Mean   :2.126         Mean   :2.191    Mean   :2.207
##  3rd Qu.:2.283      3rd Qu.:2.360         3rd Qu.:2.740    3rd Qu.:4.440
##  Max.   :5.000      Max.   :5.000         Max.   :5.000    Max.   :5.000
##                                                            NA's   :4
##  dance clubs     swimming pools      gyms            bakeries
##  Min.   :0.500   Min.   :0.50    Min.   :0.500    Min.   :0.50
##  1st Qu.:0.710   1st Qu.:0.61    1st Qu.:0.610    1st Qu.:0.61
##  Median :0.810   Median :0.76    Median :0.740    Median :0.76
##  Mean   :1.218   Mean   :1.04    Mean   :1.012    Mean   :1.20
##  3rd Qu.:1.173   3rd Qu.:0.95    3rd Qu.:0.890    3rd Qu.:0.93
##  Max.   :5.000   Max.   :5.00    Max.   :5.000    Max.   :5.00
##  NA's   :112     NA's   :479     NA's   :1017     NA's   :1046
```

```
##  beauty & spas      cafes          view points      monuments
##  Min.   :0.500   Min.   :0.500   Min.   :0.500   Min.    :0.500
##  1st Qu.:0.610   1st Qu.:0.640   1st Qu.:0.780   1st Qu.:0.840
##  Median :0.740   Median :0.800   Median :1.070   Median :1.100
##  Mean   :1.197   Mean   :1.086   Mean   :1.869   Mean    :1.621
##  3rd Qu.:0.920   3rd Qu.:1.050   3rd Qu.:2.200   3rd Qu.:1.650
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.    :5.000
##  NA's   :896     NA's   :604     NA's   :345     NA's    :303
##     gardens
##  Min.   :0.500
##  1st Qu.:0.920
##  Median :1.310
##  Mean   :1.628
##  3rd Qu.:1.690
##  Max.   :5.000
##  NA's   :225
```

df

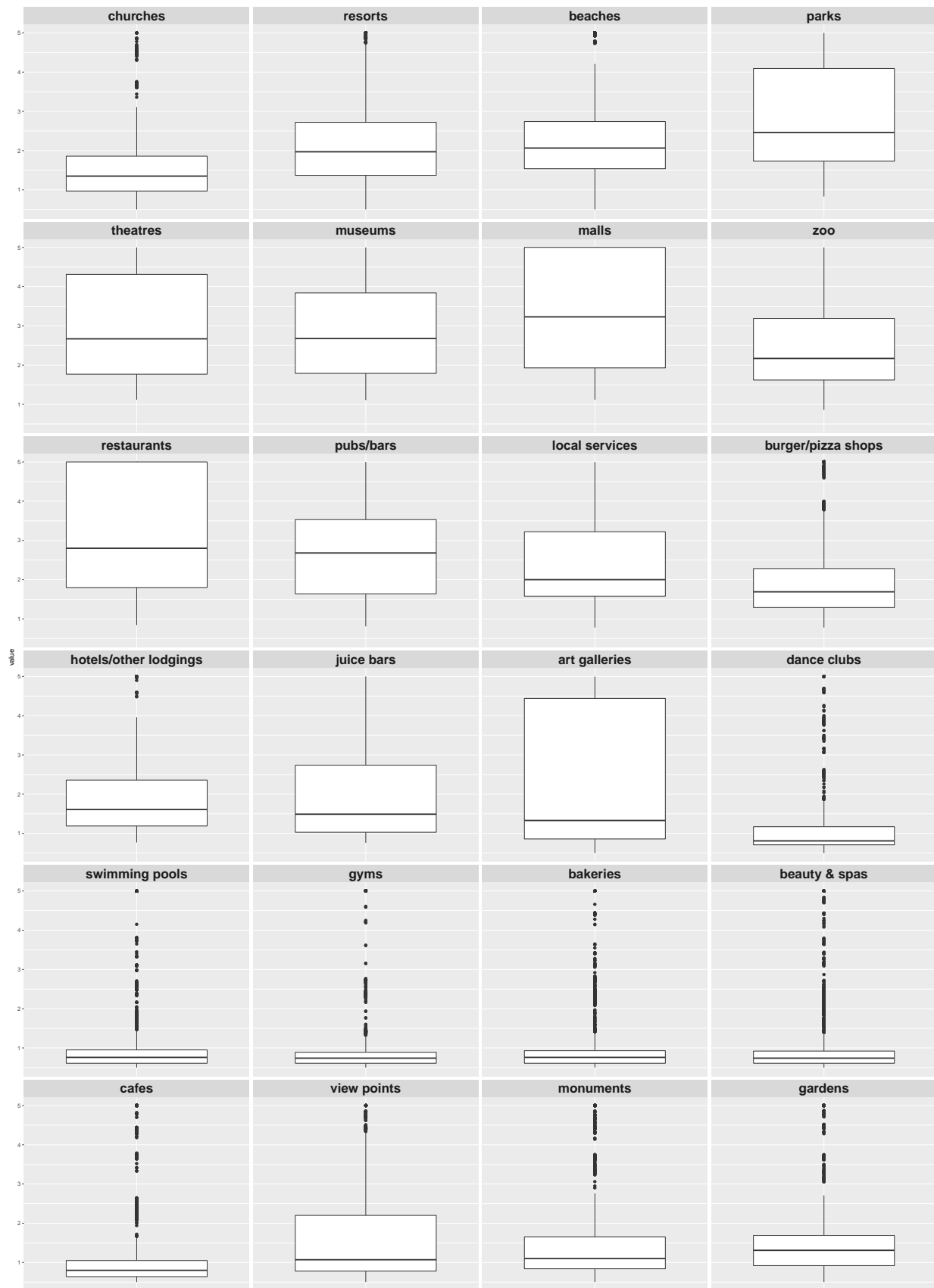```
A tibble: 5,456 x 25
User   churches resorts beaches parks theatres museums malls   zoo
<chr>   <dbl>   <dbl>   <dbl> <dbl>   <dbl>   <dbl> <dbl> <dbl>
1 User~   NA   NA   3.63 3.65   5   2.92 5   2.35
2 User~   NA   NA   3.63 3.65   5   2.92 5   2.64
3 User~   NA   NA   3.63 3.63   5   2.92 5   2.64
4 User~   NA   0.5   3.63 3.63   5   2.92 5   2.35
5 User~   NA   NA   3.63 3.63   5   2.92 5   2.64
6 User~   NA   NA   3.63 3.63   5   2.92 5   2.63
7 User~   NA   5   3.63 3.63   5   2.92 3.03   2.35
8 User~   NA   5   3.63 3.63   5   2.92 5   2.63
9 User~   NA   5   3.64 3.64   5   2.92 3.03   2.62
10 User~   NA   5   3.64 3.64   5   2.92 5   2.35
... with 5,446 more rows, and 16 more variables: restaurants <dbl>,
`pubs/bars` <dbl>, `local services` <dbl>, `burger/pizza shops` <dbl>,
`hotels/other lodgings` <dbl>, `juice bars` <dbl>, `art
galleries` <dbl>, `dance clubs` <dbl>, `swimming pools` <dbl>,
gyms <dbl>, bakeries <dbl>, `beauty & spas` <dbl>, cafes <dbl>, `view
points` <dbl>, monuments <dbl>, gardens <dbl>
```

```r
ggplot(melt(df), aes(x="", y=value)) +
  geom_boxplot() +
  facet_wrap(~variable, ncol = 4, nrow = 6) +
  theme(strip.text.x = element_text(size = 20, face = "bold"))
```

```
## Using User as id variables
```

```
## Warning: Removed 5321 rows containing non-finite values (stat_boxplot).
```

3

churches | resorts | beaches | parks

theatres | museums | malls | zoo

restaurants | pubs/bars | local services | burger/pizza shops

hotels/other lodgings | juice bars | art galleries | dance clubs

swimming pools | gyms | bakeries | beauty & spas

cafes | view points | monuments | gardens

value

x

```
corr <- cor(df[-1], use="pairwise.complete.obs")
```

```
corr
```

```
##                             churches      resorts     beaches       parks
## churches                  1.00000000  0.190703262  0.117859860  0.02471248
## resorts                   0.19070326  1.000000000  0.312876823  0.14599268
## beaches                   0.11785986  0.312876823  1.000000000  0.39599707
## parks                     0.02471248  0.145992678  0.395997073  1.00000000
## theatres                 -0.01928200  0.130839162  0.328895737  0.62686781
## museums                  -0.16339990  0.025733178  0.159925341  0.31535354
## malls                    -0.32726392 -0.075497418 -0.073862691 -0.06791422
## zoo                      -0.21986686 -0.025944989 -0.188028113 -0.12806364
## restaurants              -0.33010154 -0.059334758 -0.220620554 -0.16944248
## pubs/bars                -0.28030629 -0.090496502 -0.179141366 -0.11551513
## local services           -0.18489502 -0.210377204 -0.156981074 -0.11717256
## burger/pizza shops       -0.17403439 -0.107209483 -0.233135495 -0.16586895
## hotels/other lodgings    -0.10269830 -0.175846613 -0.180732130 -0.14573916
## juice bars               -0.17921376 -0.085145256 -0.160418204 -0.30816442
## art galleries            -0.13390620 -0.041035557 -0.131350854 -0.27090263
## dance clubs               0.05658517 -0.034249989 -0.016088036  0.02802792
## swimming pools            0.14962603 -0.048357614 -0.052000815 -0.10713685
## gyms                      0.17857342 -0.037501617 -0.089335891 -0.18483263
## bakeries                  0.19595428  0.031730918 -0.049491161 -0.19648687
## beauty & spas             0.15663534  0.038102889 -0.002389036 -0.10547119
## cafes                     0.25439095  0.003151409 -0.030970434 -0.10583775
## view points               0.30537801 -0.052725889  0.114702914  0.24596404
## monuments                 0.35014169  0.015371469  0.095132977  0.12591979
## gardens                   0.43964653  0.065037036  0.057620195  0.04894755
##                             theatres      museums       malls          zoo
## churches                 -0.019281995 -0.16339990 -0.32726392 -0.219866861
## resorts                   0.130839162  0.02573318 -0.07549742 -0.025944989
## beaches                   0.328895737  0.15992534 -0.07386269 -0.188028113
## parks                     0.626867812  0.31535354 -0.06791422 -0.128063643
## theatres                  1.000000000  0.48993680  0.07780745 -0.002594647
## museums                   0.489936801  1.00000000  0.38277405  0.200528132
## malls                     0.077807452  0.38277405  1.00000000  0.406959112
## zoo                      -0.002594647  0.20052813  0.40695911  1.000000000
## restaurants              -0.170430090  0.11471865  0.43205439  0.536870197
## pubs/bars                -0.100190723 -0.01900678  0.25656336  0.551409012
## local services           -0.124324004 -0.14832735  0.09807610  0.292452886
## burger/pizza shops       -0.104688945 -0.16365525  0.03113827  0.003473816
## hotels/other lodgings    -0.091085488 -0.13763325  0.02483324 -0.011078538
## juice bars               -0.280766976 -0.15342733  0.08910125 -0.021849821
## art galleries            -0.322425436 -0.18572724  0.09229888 -0.064045955
## dance clubs              -0.044448593 -0.14462427 -0.14285444 -0.124928621
## swimming pools           -0.134504868 -0.19860342 -0.19697966 -0.186201198
## gyms                     -0.205619728 -0.23613117 -0.21709611 -0.189113911
## bakeries                 -0.249347242 -0.24033537 -0.27682818 -0.253168566
## beauty & spas            -0.214986154 -0.24559424 -0.25903348 -0.244116893
## cafes                    -0.192254100 -0.25272386 -0.30263060 -0.288130932
## view points               0.079324183 -0.15254508 -0.41687604 -0.271414007
## monuments                 0.076210383 -0.14968588 -0.27349278 -0.183545941
## gardens                   0.054692455 -0.12541854 -0.29364237 -0.159844582
```
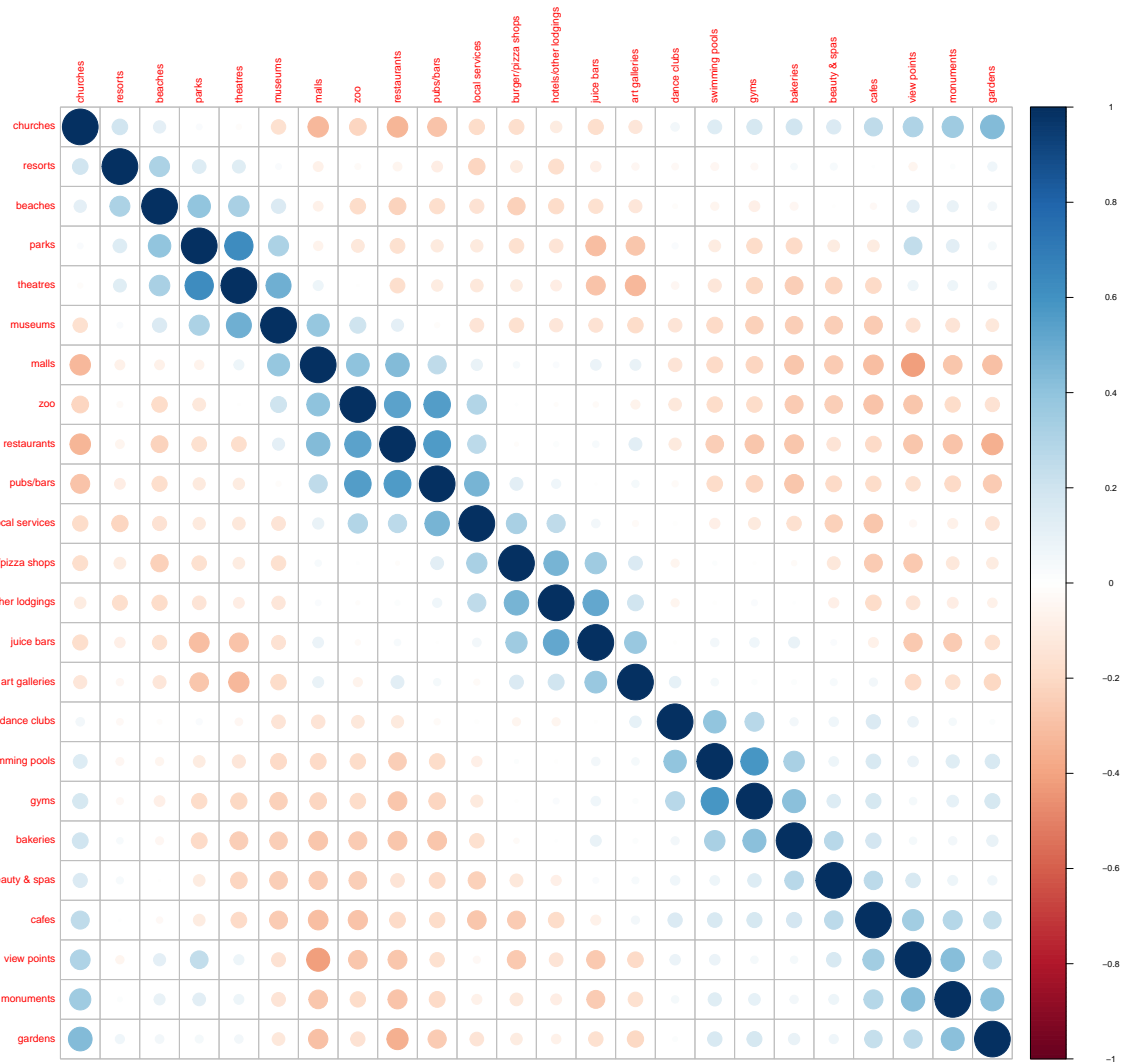
```
##                            restaurants     pubs/bars local services
## churches                   -0.33010154 -0.280306292   -0.184895019
## resorts                    -0.05933476 -0.090496502   -0.210377204
## beaches                    -0.22062055 -0.179141366   -0.156981074
## parks                      -0.16944248 -0.115515134   -0.117172559
## theatres                   -0.17043009 -0.100190723   -0.124324004
## museums                     0.11471865 -0.019006780   -0.148327352
## malls                       0.43205439  0.256563363    0.098076097
## zoo                         0.53687020  0.551409012    0.292452886
## restaurants                 1.00000000  0.564086286    0.262055861
## pubs/bars                   0.56408629  1.000000000    0.469567668
## local services             0.26205586  0.469567668    1.000000000
## burger/pizza shops         -0.01238150  0.128656647    0.321760182
## hotels/other lodgings       0.01870531  0.065379702    0.257088219
## juice bars                  0.03226012 -0.001802804    0.057389730
## art galleries               0.12869134  0.040356119   -0.027439041
## dance clubs                -0.11568882 -0.028834357   -0.008906942
## swimming pools             -0.24159684 -0.189011876   -0.083687497
## gyms                       -0.27401790 -0.216305389   -0.111664584
## bakeries                   -0.27616314 -0.275114787   -0.163202413
## beauty & spas              -0.14544524 -0.194494071   -0.232773516
## cafes                      -0.19417241 -0.183433305   -0.270710000
## view points                -0.27895182 -0.162580237   -0.038806958
## monuments                  -0.28391578 -0.195696761   -0.078330348
## gardens                    -0.35115500 -0.258182806   -0.146140363
##                            burger/pizza shops hotels/other lodgings
## churches                         -0.174034393          -0.102698301
## resorts                          -0.107209483          -0.175846613
## beaches                          -0.233135495          -0.180732130
## parks                            -0.165868945          -0.145739159
## theatres                         -0.104688945          -0.091085488
## museums                          -0.163655245          -0.137633254
## malls                             0.031138273           0.024833240
## zoo                               0.003473816          -0.011078538
## restaurants                      -0.012381498           0.018705307
## pubs/bars                         0.128656647           0.065379702
## local services                   0.321760182           0.257088219
## burger/pizza shops                1.000000000           0.469724358
## hotels/other lodgings             0.469724358           1.000000000
## juice bars                        0.354030770           0.513060734
## art galleries                     0.150169845           0.197604608
## dance clubs                      -0.050796876          -0.050849471
## swimming pools                    0.011800408          -0.005876017
## gyms                             -0.004094529           0.026135651
## bakeries                         -0.021349131           0.002793246
## beauty & spas                    -0.129701541          -0.080188886
## cafes                            -0.250413077          -0.188109390
## view points                      -0.263129314          -0.145983449
## monuments                        -0.126311400          -0.092220967
## gardens                          -0.103461156          -0.070104789
##                             juice bars art galleries  dance clubs
## churches                   -0.179213758   -0.13390620  0.056585166
## resorts                    -0.085145256   -0.04103556 -0.034249989
## beaches                    -0.160418204   -0.13135085 -0.016088036
```

```
## parks              -0.308164420  -0.27090263  0.028027921
## theatres           -0.280766976  -0.32242544 -0.044448593
## museums            -0.153427325  -0.18572724 -0.144624265
## malls               0.089101254   0.09229888 -0.142854439
## zoo                -0.021849821  -0.06404596 -0.124928621
## restaurants         0.032260117   0.12869134 -0.115688817
## pubs/bars          -0.001802804   0.04035612 -0.028834357
## local services      0.057389730  -0.02743904 -0.008906942
## burger/pizza shops  0.354030770   0.15016985 -0.050796876
## hotels/other lodgings 0.513060734  0.19760461 -0.050849471
## juice bars          1.000000000   0.37032846 -0.004583815
## art galleries       0.370328458   1.00000000  0.102728112
## dance clubs        -0.004583815   0.10272811  1.000000000
## swimming pools      0.055058092   0.04370457  0.393270734
## gyms                0.069945668   0.01822272  0.279685312
## bakeries            0.093407565   0.01708129  0.051281344
## beauty & spas       0.024223155   0.03653436  0.068860208
## cafes              -0.077415896   0.05154135  0.159670551
## view points        -0.261574971  -0.19078796  0.085472423
## monuments          -0.252268911  -0.16421909  0.042109999
## gardens            -0.163032240  -0.20862545  0.016583077
##                    swimming pools         gyms      bakeries
## churches             0.149626027  0.178573418  0.195954281
## resorts             -0.048357614 -0.037501617  0.031730918
## beaches             -0.052000815 -0.089335891 -0.049491161
## parks               -0.107136850 -0.184832633 -0.196486872
## theatres            -0.134504868 -0.205619728 -0.249347242
## museums             -0.198603424 -0.236131169 -0.240335371
## malls               -0.196979659 -0.217096111 -0.276828184
## zoo                 -0.186201198 -0.189113911 -0.253168566
## restaurants         -0.241596845 -0.274017895 -0.276163139
## pubs/bars           -0.189011876 -0.216305389 -0.275114787
## local services      -0.083687497 -0.111664584 -0.163202413
## burger/pizza shops   0.011800408 -0.004094529 -0.021349131
## hotels/other lodgings -0.005876017  0.026135651  0.002793246
## juice bars           0.055058092  0.069945668  0.093407565
## art galleries        0.043704575  0.018222722  0.017081294
## dance clubs          0.393270734  0.279685312  0.051281344
## swimming pools       1.000000000  0.580234517  0.327688668
## gyms                 0.580234517  1.000000000  0.416607402
## bakeries             0.327688668  0.416607402  1.000000000
## beauty & spas        0.072604325  0.141628438  0.271121302
## cafes                0.166632469  0.177604785  0.182133491
## view points          0.087229056  0.042538045  0.041866861
## monuments            0.135249314  0.106084702  0.054150195
## gardens              0.162300547  0.175758839  0.099409298
##                    beauty & spas        cafes view points    monuments
## churches             0.156635342  0.254390953  0.30537801  0.35014169
## resorts              0.038102889  0.003151409 -0.05272589  0.01537147
## beaches             -0.002389036 -0.030970434  0.11470291  0.09513298
## parks               -0.105471188 -0.105837752  0.24596404  0.12591979
## theatres            -0.214986154 -0.192254100  0.07932418  0.07621038
## museums             -0.245594238 -0.252723865 -0.15254508 -0.14968588
## malls               -0.259033482 -0.302630601 -0.41687604 -0.27349278
```

```
## zoo                     -0.244116893 -0.288130932 -0.27141401 -0.18354594
## restaurants             -0.145445237 -0.194172409 -0.27895182 -0.28391578
## pubs/bars               -0.194494071 -0.183433305 -0.16258024 -0.19569676
## local services          -0.232773516 -0.270710000 -0.03880696 -0.07833035
## burger/pizza shops       -0.129701541 -0.250413077 -0.26312931 -0.12631140
## hotels/other lodgings   -0.080188886 -0.188109390 -0.14598345 -0.09222097
## juice bars               0.024223155 -0.077415896 -0.26157497 -0.25226891
## art galleries            0.036534356  0.051541350 -0.19078796 -0.16421909
## dance clubs              0.068860208  0.159670551  0.08547242  0.04211000
## swimming pools           0.072604325  0.166632469  0.08722906  0.13524931
## gyms                     0.141628438  0.177604785  0.04253804  0.10608470
## bakeries                 0.271121302  0.182133491  0.04186686  0.05415020
## beauty & spas            1.000000000  0.265830662  0.16380480  0.07726239
## cafes                    0.265830662  1.000000000  0.34885979  0.28505835
## view points              0.163804803  0.348859794  1.00000000  0.42853716
## monuments                0.077262393  0.285058353  0.42853716  1.00000000
## gardens                  0.065384368  0.239669087  0.26248424  0.41193497
##                              gardens
## churches                  0.43964653
## resorts                   0.06503704
## beaches                   0.05762020
## parks                     0.04894755
## theatres                  0.05469245
## museums                  -0.12541854
## malls                    -0.29364237
## zoo                      -0.15984458
## restaurants              -0.35115500
## pubs/bars                -0.25818281
## local services           -0.14614036
## burger/pizza shops       -0.10346116
## hotels/other lodgings    -0.07010479
## juice bars               -0.16303224
## art galleries            -0.20862545
## dance clubs               0.01658308
## swimming pools            0.16230055
## gyms                      0.17575884
## bakeries                  0.09940930
## beauty & spas             0.06538437
## cafes                     0.23966909
## view points               0.26248424
## monuments                 0.41193497
## gardens                   1.00000000
```

```r
corrplot(corr)
```

```
corr %>%
  as.data.frame() %>%
  mutate(var1 = rownames(.)) %>%
  gather(var2, value, -var1) %>%
  arrange(desc(value)) %>%
  group_by(value) %>%
  dplyr::filter(row_number()==1 & var1 != var2) -> corrs

corrs %>% dplyr::filter(abs(value) > 0.3) %>% graph_from_data_frame -> corrs_graph

ggraph(corrs_graph, "linear", circular = T) +
    geom_edge_arc2(aes(edge_alpha = tan(abs(value)), edge_width = abs(value), color = value)) +
    geom_node_point(size = 2) +
    geom_node_text(aes(label = name), vjust = 1, hjust = 1) + guides(edge_alpha = "none", edge_width =
```

```
scale_edge_colour_gradientn(limits = c(-1, 1), colors = c("firebrick2", "dodgerblue2")) + theme_void(
```