Airport Connection Network Analysis

Hannah Bohan Chou

- Data set Description:

The routes & airport data are from www.openflights.org. The routes data was last updated in 2014, while the airport data was last updated in 2017.

```
#Loading the necessary libraries:
library(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
```

```
library(ggplot2)
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

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

```
library(itertools)
```

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

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(rgexf)
library(ggrepel)
library(RgoogleMaps)
library(ggmap)
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(mapproj)
```

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

```
library(sf)
```

```
## Linking to GEOS 3.8.1, GDAL 3.1.4, PROJ 6.3.1
```

```
library(OpenStreetMap)
library(devtools)
```

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

```
library(DT)
library(plyr)
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following object is masked from 'package:maps':
##
##     ozone
```

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

```
library(geosphere) # For spatial methods
library(threejs)   # threejs is used for 3-D interactive Earth Visualization
library(rworldmap) # For creating earth map
```

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

```
## ### Welcome to rworldmap ###
```

```
## For a short introduction type :   vignette('rworldmap')
```

```
library(leaflet)   # Leaflet for R provides functions to control and integrate Leaflet, a JavaScri
pt library for interactive maps, within R.
library(rgeos)     # Provides functions for handling operations on topologies.
```

```
## rgeos version: 0.5-5, (SVN revision 640)
##  GEOS runtime version: 3.8.1-CAPI-1.13.3
##  Linking to sp version: 1.4-2
##  Polygon checking: TRUE
```

```
library(raster)    # For raster image
```

```
##
## Attaching package: 'raster'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(DT)         # For creating interactive tables
library(ggplot2)
library(sp)          # For Spatial processing of data
library(ggmap)      # To reverse geocode Long/Lat
library(knitr)        # TO enable 3-D visualization embedding in the HTML page
library(rglwidget)
```

```
## The functions in the rglwidget package have been moved to rgl.
```

```
library(rgl)
```

```
##
## Attaching package: 'rgl'
```

```
## The following object is masked from 'package:rgeos':
##
##     triangulate
```

```
## The following objects are masked from 'package:threejs':
##
##     lines3d, points3d
```

```
library(sqldf)
```

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

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

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

1. Load data & Graph

1.1 Loading Data & Examine Dataframe

```
routes_url <- "https://gist.githubusercontent.com/hannahbhchou/8f79bddf4ad93a573ada0d10453fe7d5/ra
w/a3b2624b38579d0c450d76532031f3f47a269dec/routes.csv"
airport_url <- "https://gist.githubusercontent.com/hannahbhchou/5f59fb70e3d287c577af4b1d74a13cb5/r
aw/98ec7a19cbe39bd92857280fd8a02e80c9ea249f/airports.csv"

routes_df <- read.csv(routes_url, header = TRUE )
airport_df <- read.csv(airport_url, header =  TRUE)
```

```
head(routes_df)
```

```
##   airline airline.ID source.airport source.airport.id destination.airport
## 1      2B        410            AER              2965                 KZN
## 2      2B        410            ASF              2966                 KZN
## 3      2B        410            ASF              2966                 MRV
## 4      2B        410            CEK              2968                 KZN
## 5      2B        410            CEK              2968                 OVB
## 6      2B        410            DME              4029                 KZN
##   destination.airport.id codeshare stops equipment
## 1                   2990               0       CR2
## 2                   2990               0       CR2
## 3                   2962               0       CR2
## 4                   2990               0       CR2
## 5                   4078               0       CR2
## 6                   2990               0       CR2
```

```
head(airport_df)
```

```
##   Airport.ID                                        Name         City
## 1          1                              Goroka Airport       Goroka
## 2          2                              Madang Airport       Madang
## 3          3                Mount Hagen Kagamuga Airport  Mount Hagen
## 4          4                              Nadzab Airport       Nadzab
## 5          5 Port Moresby Jacksons International Airport Port Moresby
## 6          6                 Wewak International Airport        Wewak
##            Country IATA ICAO  Latitude Longtitude Altitude Timezone DST
## 1 Papua New Guinea  GKA AYGA -6.081690    145.392     5282       10   U
## 2 Papua New Guinea  MAG AYMD -5.207080    145.789       20       10   U
## 3 Papua New Guinea  HGU AYMH -5.826790    144.296     5388       10   U
## 4 Papua New Guinea  LAE AYNZ -6.569803    146.726      239       10   U
## 5 Papua New Guinea  POM AYPY -9.443380    147.220      146       10   U
## 6 Papua New Guinea  WWK AYWK -3.583830    143.669       19       10   U
##   Tz.database.time.zone    Type      Source
## 1  Pacific/Port_Moresby airport OurAirports
## 2  Pacific/Port_Moresby airport OurAirports
## 3  Pacific/Port_Moresby airport OurAirports
## 4  Pacific/Port_Moresby airport OurAirports
## 5  Pacific/Port_Moresby airport OurAirports
## 6  Pacific/Port_Moresby airport OurAirports
```

```
str(routes_df)
```

```
## 'data.frame':    67663 obs. of  9 variables:
##  $ airline               : chr  "2B" "2B" "2B" "2B" ...
##  $ airline.ID            : chr  "410" "410" "410" "410" ...
##  $ source.airport        : chr  "AER" "ASF" "ASF" "CEK" ...
##  $ source.airport.id     : chr  "2965" "2966" "2966" "2968" ...
##  $ destination.airport   : chr  "KZN" "KZN" "MRV" "KZN" ...
##  $ destination.airport.id: chr  "2990" "2990" "2962" "2990" ...
##  $ codeshare             : chr  "" "" "" "" ...
##  $ stops                 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ equipment             : chr  "CR2" "CR2" "CR2" "CR2" ...
```

```
str(airport_df)
```

```
## 'data.frame':    7698 obs. of  14 variables:
##  $ Airport.ID           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Name                 : chr  "Goroka Airport" "Madang Airport" "Mount Hagen Kagamuga Airport"
"Nadzab Airport" ...
##  $ City                 : chr  "Goroka" "Madang" "Mount Hagen" "Nadzab" ...
##  $ Country              : chr  "Papua New Guinea" "Papua New Guinea" "Papua New Guinea" "Papua
New Guinea" ...
##  $ IATA                 : chr  "GKA" "MAG" "HGU" "LAE" ...
##  $ ICAO                 : chr  "AYGA" "AYMD" "AYMH" "AYNZ" ...
##  $ Latitude             : num  -6.08 -5.21 -5.83 -6.57 -9.44 ...
##  $ Longtitude           : num  145 146 144 147 147 ...
##  $ Altitude             : int  5282 20 5388 239 146 19 112 283 165 251 ...
##  $ Timezone             : chr  "10" "10" "10" "10" ...
##  $ DST                  : chr  "U" "U" "U" "U" ...
##  $ Tz.database.time.zone: chr  "Pacific/Port_Moresby" "Pacific/Port_Moresby" "Pacific/Port_More
sby" "Pacific/Port_Moresby" ...
##  $ Type                 : chr  "airport" "airport" "airport" "airport" ...
##  $ Source               : chr  "OurAirports" "OurAirports" "OurAirports" "OurAirports" ...
```

```
#drop unnecessary columns
airport_drop_col <- c("ICAO","Altitude","Timezone","DST", "Tz.database.time.zone","Type", "Sourc
e")
routes_drop_col <- c("codeshare","stops","equipment")

airport_df <- airport_df %>% dplyr::select(-one_of(airport_drop_col))
routes_df <- routes_df %>% dplyr::select(-one_of(routes_drop_col))
```

1.2 Graph from dataframe & Graph Attributes

```
routes_edges <- routes_df %>% dplyr::select("source.airport", "destination.airport")
g <- graph_from_data_frame(d = routes_edges, directed = TRUE)
```

```
num_edge <- gsize(g)
num_vertex <- gorder(g)
print(paste("There are", num_edge, "edges."))
```

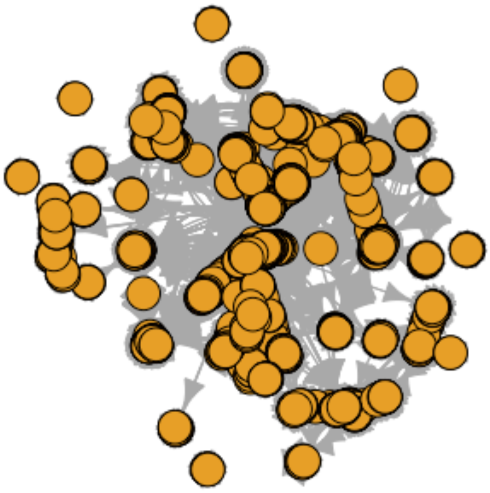
```
## [1] "There are 67663 edges."
```

```
print(paste("There are", num_vertex, "vertices."))
```

```
## [1] "There are 3425 vertices."
```

1.3 Initial Plotting

```
plot(g, vertex.label= NA, layout = layout_nicely(g))
```



Already we could see there nodes more on the outskirt, the lonely islands in terms of air traffic

2. Centrality Measures

We are using 3 centrality measures to evaluate the nodes of our graph.

```
#Run all measurements
degree_vec <- degree(g)
betweenness_vec <- betweenness(g)
closeness_vec <- closeness(g)
```

```
## Warning in closeness(g): At centrality.c:2784 :closeness centrality is not well-
## defined for disconnected graphs
```

```
in_degree_vec <- degree(g, mode = "in")
out_degree_vec <- degree(g, mode = "out")
eigen_vec <- eigen_centrality(g)$vector
```

```
#Attaching measures to the airport_df
degree_df <- as.data.frame(as.table(degree_vec))
betweenness_df <- as.data.frame(as.table(betweenness_vec))
closeness_df <- as.data.frame(as.table(closeness_vec))
in_degree_df <- as.data.frame(as.table(in_degree_vec))
out_degree_df <- as.data.frame(as.table(out_degree_vec))
eigen_df <- as.data.frame(as.table(eigen_vec))



names(degree_df)[1] <- "id"
names(degree_df)[2] <- "degree"

names(betweenness_df)[1] <- "id"
names(betweenness_df)[2] <- "betweenness"

names(closeness_df)[1] <- "id"
names(closeness_df)[2] <- "closeness"

names(in_degree_df)[1] <- "id"
names(in_degree_df)[2] <- "in_degree"

names(out_degree_df)[1] <- "id"
names(out_degree_df)[2] <- "out_degree"

names(eigen_df)[1] <- "id"
names(eigen_df)[2] <- "eigenvector"

airport_df <-  airport_df %>% left_join(degree_df, by = c("IATA" = "id"))  %>%
                left_join(in_degree_df, by = c("IATA" = "id")) %>%
                left_join(out_degree_df, by = c("IATA" = "id")) %>%
                left_join(betweenness_df, by = c("IATA" = "id")) %>%
                left_join(closeness_df, by = c("IATA" = "id")) %>%
                left_join(eigen_df, by = c("IATA" = "id"))

airport_df <- airport_df[complete.cases(airport_df),]
```

2.1 Degree Centrality

2.1.1 Maximum & Minimum Degree

```
max_degree <- max(degree_vec)
min_degree <- min(degree_vec)
print(paste("Maximum degree is", max_degree, "degree."))
```

```
## [1] "Maximum degree is 1826 degree."
```

```
print(paste("Minimum degree is", min_degree, "degree."))
```

```
## [1] "Minimum degree is 1 degree."
```

```
max_in_degree <- max(in_degree_vec)
min_in_degree <- min(in_degree_vec)
print(paste("Maximum in degree is", max_in_degree, "degree, which means this airport receives flig
hts  from", max_in_degree, "destinations."))
```

```
## [1] "Maximum in degree is 911 degree, which means this airport receives flights  from 911 desti
nations."
```

```
print(paste("Minimum degree is", min_in_degree, "degree, which means this airport doesn't receive
any flights."))
```

```
## [1] "Minimum degree is 0 degree, which means this airport doesn't receive any flights."
```

```
max_out_degree <- max(out_degree_vec)
min_out_degree <- min(out_degree_vec)
print(paste("Maximum out degree is", max_out_degree, "degree, which means this airport receives fl
ights  from", max_out_degree, "destinations."))
```

```
## [1] "Maximum out degree is 915 degree, which means this airport receives flights  from 915 dest
inations."
```

```
print(paste("Minimum degree is", min_out_degree, "degree, which means this airport  doesn't have d
eparting flights."))
```

```
## [1] "Minimum degree is 0 degree, which means this airport  doesn't have departing flights."
```

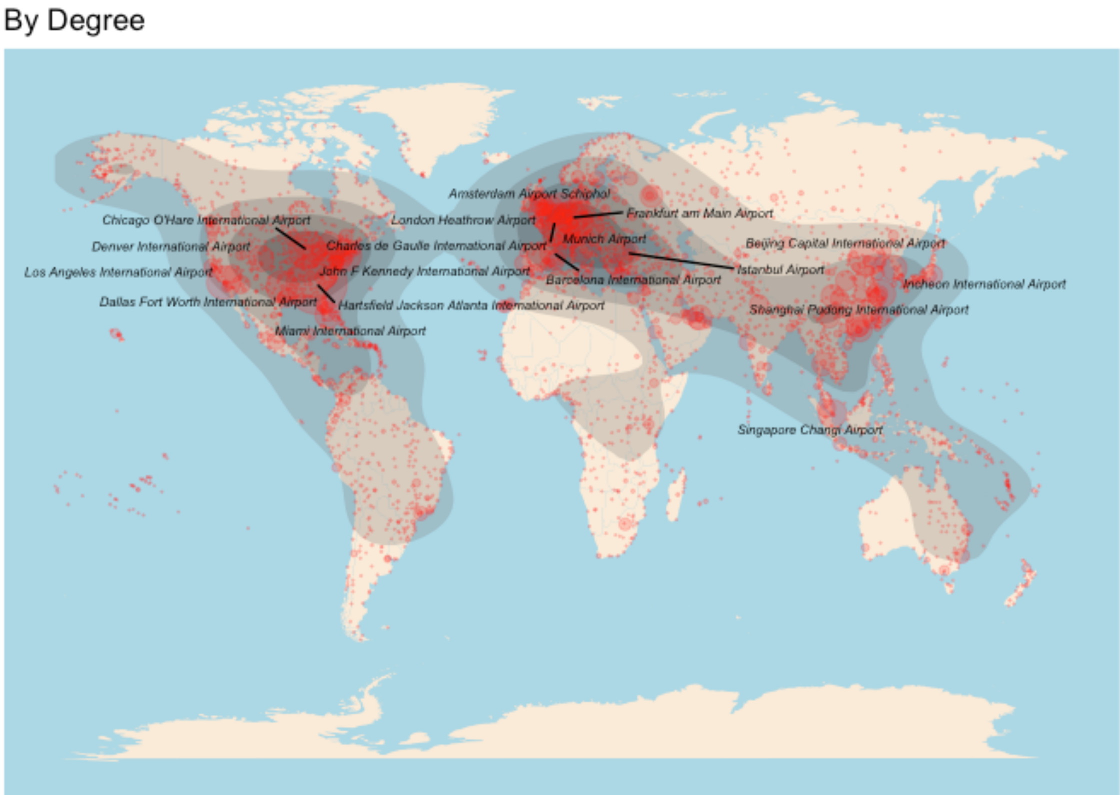
2.1.2 Top 20 Degree Airport

```
top20_degree_df <- airport_df[order(airport_df$degree, decreasing = TRUE),][1:20,c("IATA", "Name",
"Country", "City","degree")]
top20_degree_df
```

```
##      IATA                                      Name              Country
## 3483  ATL Hartsfield Jackson Atlanta International Airport        United States
## 3631  ORD            Chicago O'Hare International Airport        United States
## 3171  PEK           Beijing Capital International Airport                China
## 503   LHR                         London Heathrow Airport       United Kingdom
## 1347  CDG          Charles de Gaulle International Airport               France
## 337   FRA                       Frankfurt am Main Airport              Germany
## 3286  LAX               Los Angeles International Airport        United States
## 3471  DFW          Dallas Fort Worth International Airport        United States
## 3598  JFK            John F Kennedy International Airport        United States
## 575   AMS                    Amsterdam Airport Schiphol          Netherlands
## 3208  PVG           Shanghai Pudong International Airport                China
## 3125  SIN                        Singapore Changi Airport            Singapore
## 1187  BCN                 Barcelona International Airport                Spain
## 3726  ICN                   Incheon International Airport          South Korea
## 3552  DEN                    Denver International Airport        United States
## 3377  MIA                     Miami International Airport        United States
## 343   MUC                                  Munich Airport              Germany
## 7630  IST                                Istanbul Airport               Turkey
## 2101  DXB                     Dubai International Airport United Arab Emirates
## 2916  HKG                 Hong Kong International Airport            Hong Kong
##                  City degree
## 3483          Atlanta   1826
## 3631          Chicago   1108
## 3171          Beijing   1069
## 503            London   1051
## 1347            Paris   1041
## 337         Frankfurt    990
## 3286      Los Angeles    990
## 3471 Dallas-Fort Worth    936
## 3598         New York    911
## 575         Amsterdam    903
## 3208         Shanghai    825
## 3125        Singapore    820
## 1187        Barcelona    783
## 3726            Seoul    740
## 3552           Denver    735
## 3377            Miami    734
## 343            Munich    728
## 7630         Istanbul    719
## 2101            Dubai    710
## 2916        Hong Kong    710
```

```
#set ggplot theme
world_theme <- theme(panel.background = element_rect(fill = "lightblue",
               colour = "lightblue"),
               panel.grid.major = element_blank(),
               panel.grid.minor = element_blank(),
               # surpress legend
               legend.position = "none",
               axis.line=element_blank(),
               axis.text.x=element_blank(),
               axis.text.y=element_blank(),
               axis.ticks=element_blank(),
               axis.title.x=element_blank(),
               axis.title.y=element_blank())
```

```
#set lower bound for label showing
thres <-top20_degree_df[20, "degree"]

degree_plot <- ggplot(airport_df, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  stat_density2d(aes(fill = ..level..,  alpha = I(.3)),
                 size = 1, bins = 5, data = airport_df,
                 geom = "polygon") +
  geom_point(color="red", alpha = .2, size=airport_df$degree/150) +
  # define color of density polygons
  scale_fill_gradient(low = "grey50", high = "grey20") +
  world_theme +
  geom_text_repel(data = subset(airport_df, degree > thres), aes(x=Longtitude, y= Latitude, label=
Name), color = "black", fontface = "italic", size = 2, max.overlaps = Inf) + ggtitle("By Degree")

degree_plot
```




```
top20_in_degree_df <- airport_df[order(airport_df$in_degree, decreasing = TRUE),][1:20,c("IATA", "
Name", "Country", "City","in_degree")]
top20_in_degree_df
```

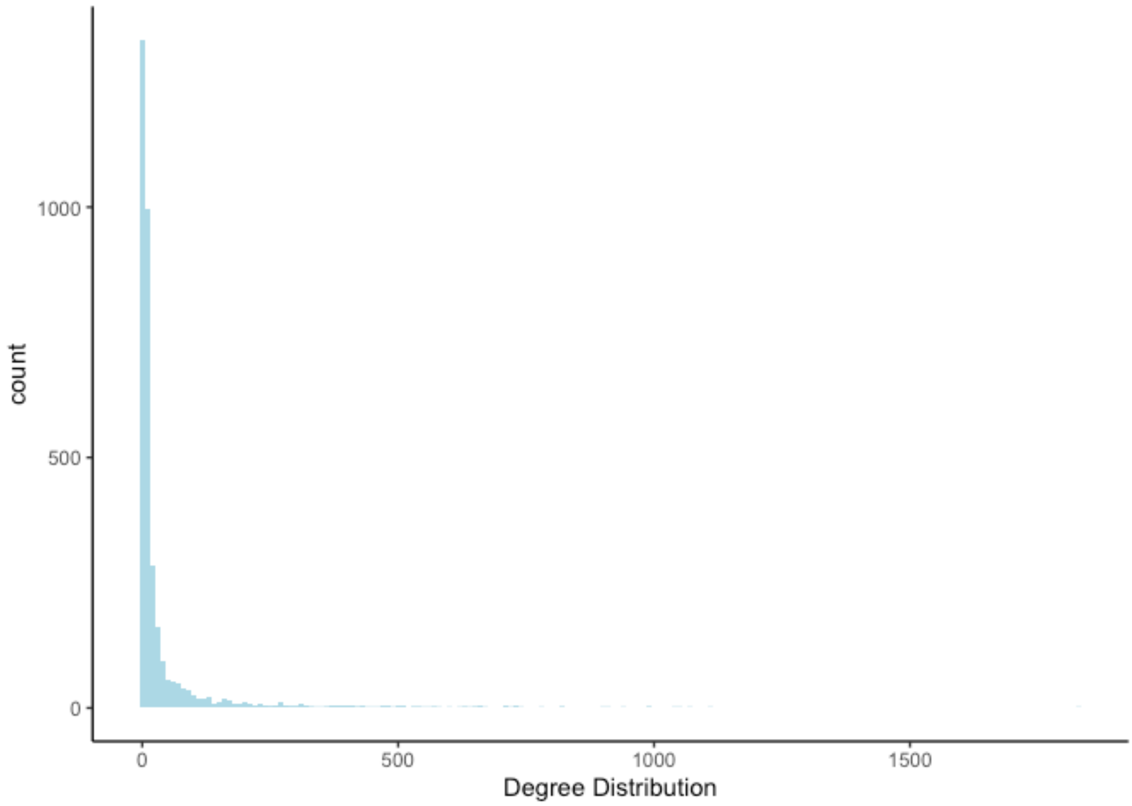
```
##      IATA                                         Name             Country
## 3483  ATL Hartsfield Jackson Atlanta International Airport       United States
## 3631  ORD             Chicago O'Hare International Airport       United States
## 3171  PEK            Beijing Capital International Airport               China
## 503   LHR                           London Heathrow Airport      United Kingdom
## 1347  CDG          Charles de Gaulle International Airport              France
## 3286  LAX              Los Angeles International Airport       United States
## 337   FRA                         Frankfurt am Main Airport             Germany
## 3471  DFW          Dallas Fort Worth International Airport       United States
## 3598  JFK             John F Kennedy International Airport       United States
## 575   AMS                        Amsterdam Airport Schiphol         Netherlands
## 3208  PVG            Shanghai Pudong International Airport               China
## 3125  SIN                          Singapore Changi Airport           Singapore
## 1187  BCN                Barcelona International Airport               Spain
## 3552  DEN                   Denver International Airport       United States
## 3726  ICN                  Incheon International Airport         South Korea
## 3377  MIA                    Miami International Airport       United States
## 7630  IST                               Istanbul Airport              Turkey
## 343   MUC                                 Munich Airport             Germany
## 2916  HKG                Hong Kong International Airport           Hong Kong
## 2101  DXB                    Dubai International Airport United Arab Emirates
##                  City in_degree
## 3483          Atlanta       911
## 3631          Chicago       550
## 3171          Beijing       534
## 503            London       524
## 1347            Paris       517
## 3286      Los Angeles       498
## 337         Frankfurt       493
## 3471 Dallas-Fort Worth       467
## 3598         New York       455
## 575         Amsterdam       450
## 3208         Shanghai       414
## 3125        Singapore       412
## 1187        Barcelona       392
## 3552           Denver       374
## 3726            Seoul       370
## 3377            Miami       366
## 7630         Istanbul       361
## 343            Munich       360
## 2916        Hong Kong       355
## 2101            Dubai       354
```

```
top20_out_degree_df <- airport_df[order(airport_df$out_degree, decreasing = TRUE),][1:20,c("IATA",
"Name", "Country", "City","out_degree")]
top20_out_degree_df
```

```
##      IATA                                          Name              Country
## 3483  ATL Hartsfield Jackson Atlanta International Airport        United States
## 3631  ORD            Chicago O'Hare International Airport        United States
## 3171  PEK           Beijing Capital International Airport                China
## 503   LHR                          London Heathrow Airport       United Kingdom
## 1347  CDG          Charles de Gaulle International Airport               France
## 337   FRA                           Frankfurt am Main Airport              Germany
## 3286  LAX             Los Angeles International Airport        United States
## 3471  DFW          Dallas Fort Worth International Airport        United States
## 3598  JFK            John F Kennedy International Airport        United States
## 575   AMS                      Amsterdam Airport Schiphol          Netherlands
## 3208  PVG           Shanghai Pudong International Airport                China
## 3125  SIN                         Singapore Changi Airport            Singapore
## 1187  BCN                Barcelona International Airport                Spain
## 3726  ICN                  Incheon International Airport          South Korea
## 343   MUC                                  Munich Airport              Germany
## 3377  MIA                    Miami International Airport        United States
## 3552  DEN                   Denver International Airport        United States
## 7630  IST                                Istanbul Airport               Turkey
## 498   LGW                           London Gatwick Airport       United Kingdom
## 2101  DXB                    Dubai International Airport United Arab Emirates
##                  City out_degree
## 3483          Atlanta        915
## 3631          Chicago        558
## 3171          Beijing        535
## 503            London        527
## 1347            Paris        524
## 337         Frankfurt        497
## 3286      Los Angeles        492
## 3471 Dallas-Fort Worth        469
## 3598         New York        456
## 575         Amsterdam        453
## 3208         Shanghai        411
## 3125        Singapore        408
## 1187        Barcelona        391
## 3726            Seoul        370
## 343            Munich        368
## 3377            Miami        368
## 3552           Denver        361
## 7630         Istanbul        358
## 498            London        356
## 2101            Dubai        356
```

2.1.3 Degree Histogram & Statistics

```
degree_hist <- ggplot(degree_df,aes(x= degree)) +
               geom_histogram(binwidth = 10, fill = "lightblue") +
               xlab(label = "Degree Distribution") +
               theme_classic()
degree_hist
```




```
psych::describe(degree_df$degree)
```

```
##    vars    n  mean     sd median trimmed mad min  max range skew kurtosis   se
## X1    1 3425 39.51 106.72      8   14.63 8.9   1 1826  1825 6.03    51.73 1.82
```

We could see we have a very right-skewed distribution, as most of the airports have small number of degree, while the top tiers have plenty.

Who are the medians?

```
eightdegree_df <- airport_df[which(airport_df$degree==8),c("IATA", "Name", "Country", "City","degr
ee")]

sample_n(eightdegree_df, 20)
```

```
##    IATA                                          Name          Country
## 1   DAU                                 Daru Airport Papua New Guinea
## 2   YNG              Youngstown Warren Regional Airport    United States
## 3   PZI                              Bao'anying Airport            China
## 4   UUA                                 Bugulma Airport           Russia
## 5   YYH                                Taloyoak Airport           Canada
## 6   YHO                                Hopedale Airport           Canada
## 7   BYC                                 Yacuiba Airport          Bolivia
## 8   NHV                               Nuku Hiva Airport French Polynesia
## 9   OUL                                    Oulu Airport          Finland
## 10  ABA                                  Abakan Airport           Russia
## 11  AVA                       Anshun Huangguoshu Airport            China
## 12  HUH                            Huahine-Fare Airport French Polynesia
## 13  YBG                                  CFB Bagotville           Canada
## 14  YUT                             Repulse Bay Airport           Canada
## 15  COQ                              Choibalsan Airport         Mongolia
## 16  FKS                               Fukushima Airport            Japan
## 17  CXI                    Cassidy International Airport         Kiribati
## 18  KWJ                                 Gwangju Airport      South Korea
## 19  SBY   Salisbury Ocean City Wicomico Regional Airport    United States
## 20  YXC Cranbrook/Canadian Rockies International Airport           Canada
##             City degree
## 1           Daru      8
## 2     Youngstown      8
## 3      Panzhihua      8
## 4        Bugulma      8
## 5     Spence Bay      8
## 6       Hopedale      8
## 7        Yacuiba      8
## 8      Nuku Hiva      8
## 9           Oulu      8
## 10        Abakan      8
## 11        Anshun      8
## 12 Huahine Island      8
## 13    Bagotville      8
## 14   Repulse Bay      8
## 15    Choibalsan      8
## 16     Fukushima      8
## 17    Kiritimati      8
## 18       Kwangju      8
## 19     Salisbury      8
## 20     Cranbrook      8
```

These are mostly regional airport which travel to and from 4 other airports.

2.1.4 In Degree & Out Degree Difference

```
airport_df$degree_diff <- with(airport_df, out_degree - in_degree)
```

```
most_outgoing <- airport_df[order(airport_df$degree_diff, decreasing = TRUE),][1:20,]
most_outgoing[,c("IATA", "Name", "Country", "City","in_degree", "out_degree")]
```

```
##      IATA                                                  Name
## 2003  JED                     King Abdulaziz International Airport
## 3367  HOU                                 William P Hobby Airport
## 343   MUC                                          Munich Airport
## 3129  BNE                          Brisbane International Airport
## 3631  ORD                     Chicago O'Hare International Airport
## 1318  MRS                               Marseille Provence Airport
## 1347  CDG                  Charles de Gaulle International Airport
## 3479  STL                   St Louis Lambert International Airport
## 3548  MDW                     Chicago Midway International Airport
## 3168  SYD             Sydney Kingsford Smith International Airport
## 3515  IAD                  Washington Dulles International Airport
## 3659  MSP Minneapolis-St Paul International/Wold-Chamberlain Airport
## 3896  JIB                                Djibouti-Ambouli Airport
## 474   MAN                                      Manchester Airport
## 1078  SID                     Amílcar Cabral International Airport
## 2093  AUH                        Abu Dhabi International Airport
## 3147  MEL                        Melbourne International Airport
## 3260  MCI                      Kansas City International Airport
## 3332  ADQ                                          Kodiak Airport
## 3518  MKE                   General Mitchell International Airport
##                   Country           City in_degree out_degree
## 2003         Saudi Arabia         Jeddah       183        194
## 3367        United States        Houston        70         79
## 343               Germany         Munich       360        368
## 3129            Australia       Brisbane       144        152
## 3631        United States        Chicago       550        558
## 1318               France      Marseille       129        136
## 1347               France          Paris       517        524
## 3479        United States      St. Louis       107        114
## 3548        United States        Chicago       132        139
## 3168            Australia         Sydney       202        208
## 3515        United States     Washington       190        196
## 3659        United States    Minneapolis       212        218
## 3896             Djibouti       Djibouti        17         23
## 474        United Kingdom     Manchester       311        316
## 1078           Cape Verde Amilcar Cabral        15         20
## 2093 United Arab Emirates      Abu Dhabi       236        241
## 3147            Australia      Melbourne       132        137
## 3260        United States    Kansas City        77         82
## 3332        United States         Kodiak         6         11
## 3518        United States      Milwaukee        60         65
```

```
out_going_plot <- ggplot(most_outgoing, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  world_theme +
  geom_point(color="red", alpha = .2, size=most_outgoing$degree_diff) +
  geom_text_repel(data = most_outgoing, (aes(x=Longtitude, y= Latitude, label=Name)), color = "bla
ck", fontface = "italic", size = 2, max.overlaps = Inf) +
  ggtitle("Most Out Going Airport")

out_going_plot
```

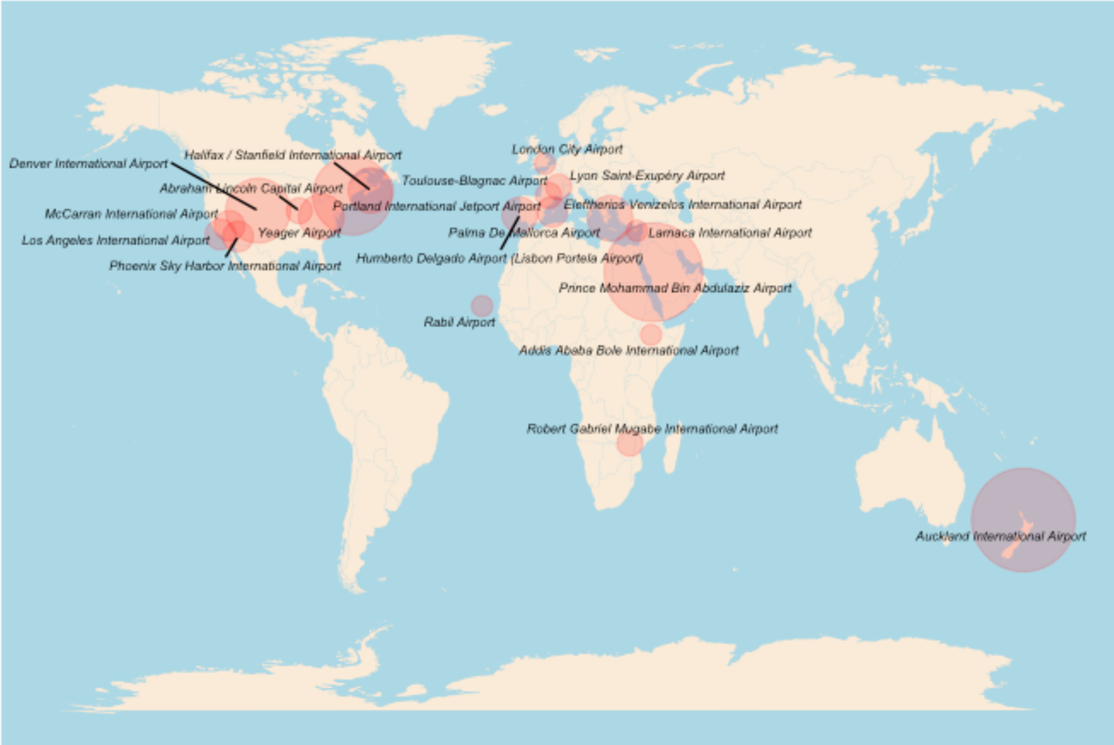
Most Out Going Airport


```
most_incoming <- airport_df[order(airport_df$degree_diff, decreasing = FALSE),][1:20,]
most_incoming[,c("IATA", "Name", "Country", "City","in_degree", "out_degree")]
```

```
##      IATA                                            Name        Country
## 1937  AKL                   Auckland International Airport    New Zealand
## 2005  MED            Prince Mohammad Bin Abdulaziz Airport   Saudi Arabia
## 3663  PWM           Portland International Jetport Airport  United States
## 3552  DEN                     Denver International Airport  United States
## 73    YHZ         Halifax / Stanfield International Airport         Canada
## 3736  ATH       Eleftherios Venizelos International Airport         Greece
## 4027  CRW                                   Yeager Airport  United States
## 1596  LIS Humberto Delgado Airport (Lisbon Portela Airport)       Portugal
## 3264  PHX          Phoenix Sky Harbor International Airport  United States
## 3286  LAX                Los Angeles International Airport  United States
## 3678  LAS                   McCarran International Airport  United States
## 3788  PMI                       Palma De Mallorca Airport          Spain
## 983   HRE       Robert Gabriel Mugabe International Airport       Zimbabwe
## 1239  TLS                         Toulouse-Blagnac Airport         France
## 1300  LYS                       Lyon Saint-Exupéry Airport         France
## 4069  SPI                  Abraham Lincoln Capital Airport  United States
## 499   LCY                       London City Airport United Kingdom
## 1079  BVC                                    Rabil Airport     Cape Verde
## 1083  ADD            Addis Ababa Bole International Airport       Ethiopia
## 1167  LCA                    Larnaca International Airport         Cyprus
##                   City in_degree out_degree
## 1937          Auckland       117         96
## 2005           Madinah        59         39
## 3663          Portland        18          2
## 3552            Denver       374        361
## 73             Halifax        52         43
## 3736            Athens       206        197
## 4027        Charleston        15          6
## 1596            Lisbon       221        214
## 3264           Phoenix       257        251
## 3286       Los Angeles       498        492
## 3678         Las Vegas       252        246
## 3788 Palma de Mallorca       277        271
## 983             Harare        31         26
## 1239          Toulouse        83         78
## 1300              Lyon       140        135
## 4069       Springfield         5          0
## 499             London        66         62
## 1079         Boa Vista        16         12
## 1083       Addis Ababa       109        105
## 1167           Larnaca        97         93
```

```
in_coming_plot <- ggplot(most_incoming, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  world_theme +
  geom_point(color="red", alpha = .2, size=abs(most_incoming$degree_diff)) +
  geom_text_repel(data = most_incoming, (aes(x=Longtitude, y= Latitude, label=Name)), color = "bla
ck", fontface = "italic", size = 2, max.overlaps = Inf) +
  ggtitle("Most In Coming Airport")

in_coming_plot
```

Most In Coming Airport



Interestingly, the two Saudi airports Prince Mohammad Bin Abdulaziz Airport in Madinah and King Abdulaziz International Airport in Jeddah, both ranked high for the degree differences, one for incoming one for out going. It may suggest that a lot of people are visiting Saudi Arabia by entering Madinah and leaving through Jeddah, that's why more routes are accommodating such needs.

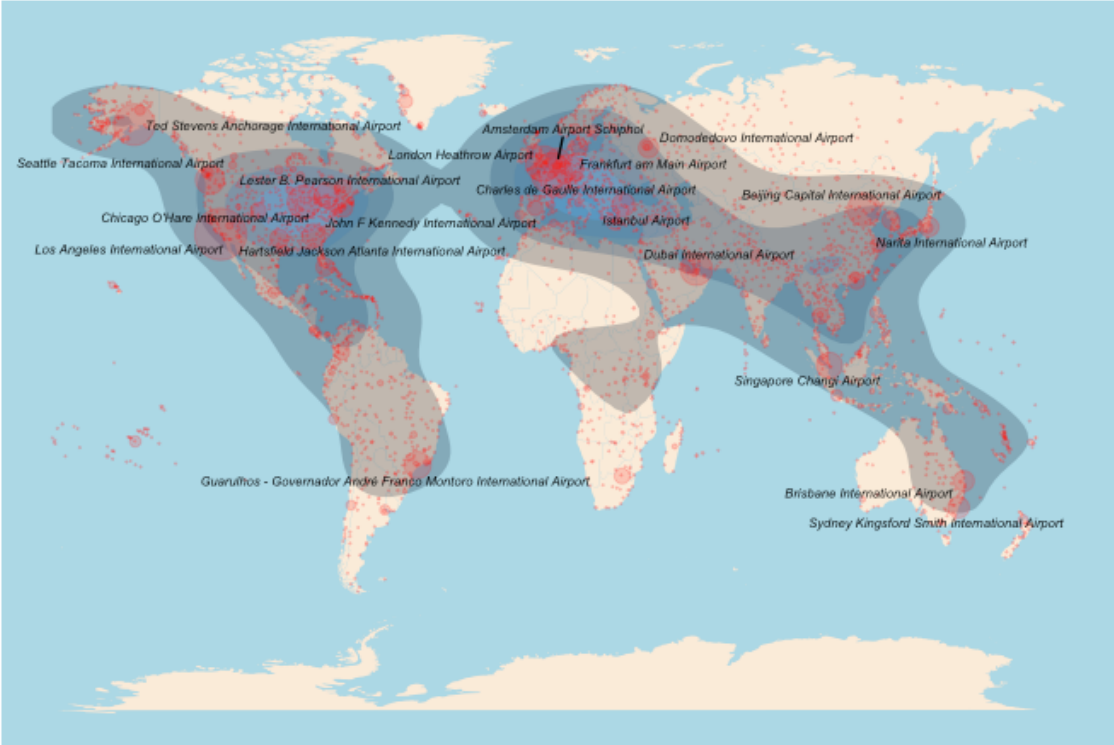
2.2 Betweenness Centrality

```
top20_betweenness_df <- airport_df[order(airport_df$betweenness, decreasing = TRUE),][1:20,]
top20_betweenness_df[,c("IATA", "Name", "Country", "City","betweenness")]
```

```
##      IATA                                                          Name
## 3286  LAX                              Los Angeles International Airport
## 3575  ANC                        Ted Stevens Anchorage International Airport
## 1347  CDG                          Charles de Gaulle International Airport
## 503   LHR                                         London Heathrow Airport
## 3631  ORD                             Chicago O'Hare International Airport
## 3171  PEK                            Beijing Capital International Airport
## 2101  DXB                                     Dubai International Airport
## 337   FRA                                       Frankfurt am Main Airport
## 3378  SEA                             Seattle Tacoma International Airport
## 2437  GRU Guarulhos - Governador André Franco Montoro International Airport
## 3125  SIN                                         Singapore Changi Airport
## 192   YYZ                          Lester B. Pearson International Airport
## 575   AMS                                       Amsterdam Airport Schiphol
## 3483  ATL                 Hartsfield Jackson Atlanta International Airport
## 7630  IST                                                 Istanbul Airport
## 3168  SYD                       Sydney Kingsford Smith International Airport
## 3129  BNE                                   Brisbane International Airport
## 3816  DME                                 Domodedovo International Airport
## 3598  JFK                              John F Kennedy International Airport
## 2182  NRT                                     Narita International Airport
##                   Country        City betweenness
## 3286        United States Los Angeles   1034522.4
## 3575        United States   Anchorage    820399.3
## 1347               France       Paris    813854.2
## 503        United Kingdom      London    702368.6
## 3631        United States     Chicago    664992.4
## 3171                China     Beijing    651405.4
## 2101 United Arab Emirates       Dubai    634412.5
## 337               Germany   Frankfurt    587555.3
## 3378        United States     Seattle    566562.7
## 2437               Brazil   Sao Paulo    521839.4
## 3125            Singapore   Singapore    504163.9
## 192                Canada     Toronto    482539.9
## 575           Netherlands   Amsterdam    460926.9
## 3483        United States     Atlanta    447437.6
## 7630               Turkey    Istanbul    442873.1
## 3168            Australia      Sydney    407827.9
## 3129            Australia    Brisbane    392096.6
## 3816               Russia      Moscow    377396.6
## 3598        United States    New York    375816.7
## 2182                Japan       Tokyo    369420.6
```

```
#set lower bound for label showing
thres <-top20_betweenness_df[20, "betweenness"]

betweenness_plot <- ggplot(airport_df, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  stat_density2d(aes(fill = ..level..,  alpha = I(.3)),
                 size = 1, bins = 5, data = airport_df,
                 geom = "polygon") +
  geom_point(color="red", alpha = .2, size=airport_df$betweenness/100000) +
  world_theme +
geom_text_repel(data = subset(airport_df, betweenness>= thres), aes(x=Longtitude, y= Latitude, lab
el=Name), color = "black", fontface = "italic", size = 2, max.overlaps = Inf) +
  ggtitle("By Betweenness")

betweenness_plot
```

By Betweenness



Which airports are Top Betweenness but not Top Degree?

```
`%nin%` = Negate(`%in%`)

for (i in top20_betweenness_df$Name){
  if (i %nin% top20_degree_df$Name){
    print(i)
  }
}
```

```
## [1] "Ted Stevens Anchorage International Airport"
## [1] "Seattle Tacoma International Airport"
## [1] "Guarulhos - Governador André Franco Montoro International Airport"
## [1] "Lester B. Pearson International Airport"
## [1] "Sydney Kingsford Smith International Airport"
## [1] "Brisbane International Airport"
## [1] "Domodedovo International Airport"
## [1] "Narita International Airport"
```

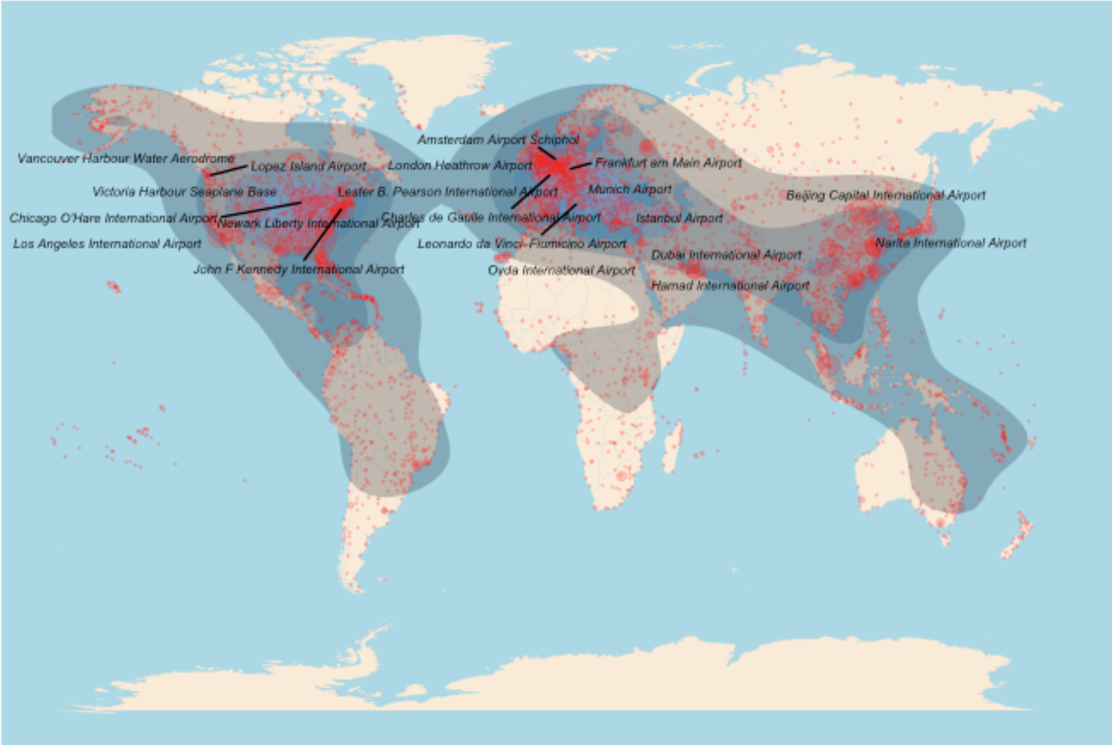
2.3 Closeness Centrality

```
top20_closeness_df <- airport_df[order(airport_df$closeness, decreasing = TRUE),][1:20,]
top20_closeness_df[,c("IATA", "Name", "Country", "City","closeness")]
```

```
##      IATA                                      Name              Country
## 3884  YWH          Victoria Harbour Seaplane Base               Canada
## 4207  CXH       Vancouver Harbour Water Aerodrome               Canada
## 4771  LPS                     Lopez Island Airport        United States
## 337   FRA                Frankfurt am Main Airport              Germany
## 1561  VDA                Ovda International Airport               Israel
## 1347  CDG Charles de Gaulle International Airport               France
## 503   LHR                  London Heathrow Airport       United Kingdom
## 2101  DXB               Dubai International Airport United Arab Emirates
## 575   AMS                Amsterdam Airport Schiphol          Netherlands
## 3286  LAX       Los Angeles International Airport        United States
## 3598  JFK    John F Kennedy International Airport        United States
## 192   YYZ Lester B. Pearson International Airport               Canada
## 7630  IST                         Istanbul Airport               Turkey
## 3631  ORD    Chicago O'Hare International Airport        United States
## 343   MUC                           Munich Airport              Germany
## 3171  PEK   Beijing Capital International Airport                China
## 2182  NRT              Narita International Airport                Japan
## 1515  FCO     Leonardo da Vinci–Fiumicino Airport                Italy
## 3296  EWR    Newark Liberty International Airport        United States
## 6828  DOH               Hamad International Airport                Qatar
##             City    closeness
## 3884    Victoria 6.673785e-06
## 4207   Vancouver 6.526393e-06
## 4771       Lopez 6.121525e-06
## 337    Frankfurt 5.901794e-06
## 1561        Ovda 5.901550e-06
## 1347       Paris 5.899914e-06
## 503       London 5.898731e-06
## 2101       Dubai 5.895079e-06
## 575    Amsterdam 5.894315e-06
## 3286 Los Angeles 5.892092e-06
## 3598    New York 5.890496e-06
## 192      Toronto 5.886959e-06
## 7630    Istanbul 5.884915e-06
## 3631     Chicago 5.884603e-06
## 343       Munich 5.884118e-06
## 3171     Beijing 5.884084e-06
## 2182       Tokyo 5.881626e-06
## 1515        Rome 5.881592e-06
## 3296      Newark 5.880796e-06
## 6828        Doha 5.880796e-06
```

```
#set lower bound for label showing
thres <-top20_closeness_df[20, "closeness"]

closeness_plot <- ggplot(airport_df, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  stat_density2d(aes(fill = ..level..,  alpha = I(.3)),
                 size = 1, bins = 5, data = airport_df,
                 geom = "polygon") +
  geom_point(color="red", alpha = .2, size=airport_df$degree/200) +
  world_theme +
geom_text_repel(data = subset(airport_df, closeness >= thres), aes(x=Longtitude, y= Latitude, labe
l=Name), color = "black", fontface = "italic", size = 2, max.overlaps = Inf) +
  ggtitle("By Closeness")

closeness_plot
```

By Closeness




```
for (i in top20_closeness_df$Name){
  if (i %nin% top20_degree_df$Name){
    print(i)
  }
}
```

```
## [1] "Victoria Harbour Seaplane Base"
## [1] "Vancouver Harbour Water Aerodrome"
## [1] "Lopez Island Airport"
## [1] "Ovda International Airport"
## [1] "Lester B. Pearson International Airport"
## [1] "Narita International Airport"
## [1] "Leonardo da Vinci–Fiumicino Airport"
## [1] "Newark Liberty International Airport"
## [1] "Hamad International Airport"
```

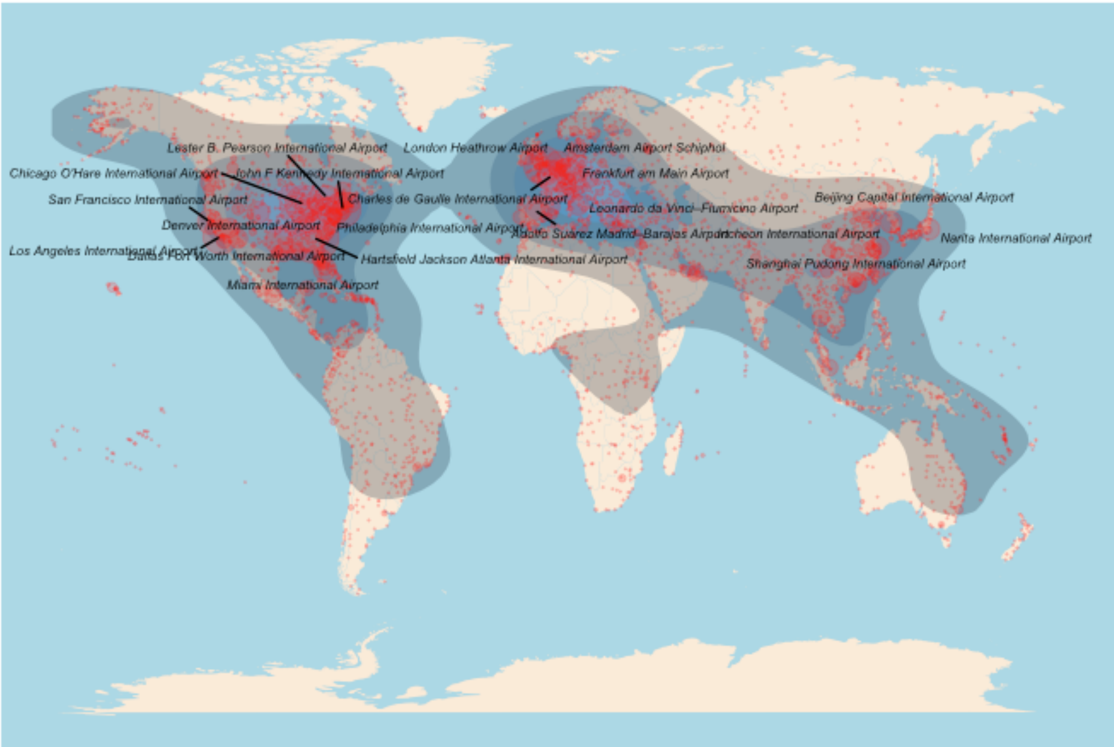
2.4 Eigenvector Centrality

```
top20_eigen_df <- airport_df[order(airport_df$eigenvector, decreasing = TRUE),][1:20,]
top20_eigen_df[,c("IATA", "Name", "Country", "City","eigenvector")]
```

```
##      IATA                                            Name        Country
## 3483  ATL Hartsfield Jackson Atlanta International Airport  United States
## 503   LHR                          London Heathrow Airport United Kingdom
## 3631  ORD             Chicago O'Hare International Airport  United States
## 3598  JFK             John F Kennedy International Airport  United States
## 3286  LAX                Los Angeles International Airport  United States
## 1347  CDG          Charles de Gaulle International Airport         France
## 3471  DFW          Dallas Fort Worth International Airport  United States
## 337   FRA                      Frankfurt am Main Airport        Germany
## 3271  SFO              San Francisco International Airport  United States
## 192   YYZ          Lester B. Pearson International Airport         Canada
## 575   AMS                      Amsterdam Airport Schiphol    Netherlands
## 3171  PEK            Beijing Capital International Airport          China
## 3377  MIA                      Miami International Airport  United States
## 3552  DEN                     Denver International Airport  United States
## 3208  PVG            Shanghai Pudong International Airport          China
## 3726  ICN                    Incheon International Airport    South Korea
## 2182  NRT                     Narita International Airport          Japan
## 1515  FCO              Leonardo da Vinci–Fiumicino Airport          Italy
## 1197  MAD             Adolfo Suárez Madrid–Barajas Airport          Spain
## 3553  PHL               Philadelphia International Airport  United States
##                  City eigenvector
## 3483          Atlanta   1.0000000
## 503            London   0.7704645
## 3631          Chicago   0.7442810
## 3598         New York   0.7064476
## 3286      Los Angeles   0.6884858
## 1347            Paris   0.5834824
## 3471 Dallas-Fort Worth   0.5284687
## 337         Frankfurt   0.5272327
## 3271    San Francisco   0.4687981
## 192           Toronto   0.4573527
## 575         Amsterdam   0.4411905
## 3171          Beijing   0.4376281
## 3377            Miami   0.4367697
## 3552           Denver   0.4270043
## 3208         Shanghai   0.4084017
## 3726            Seoul   0.4081232
## 2182            Tokyo   0.3996506
## 1515             Rome   0.3898507
## 1197           Madrid   0.3896220
## 3553     Philadelphia   0.3896203
```

```
#set lower bound for label showing
thres <-top20_eigen_df[20, "eigenvector"]

eigen_plot <- ggplot(airport_df, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  stat_density2d(aes(fill = ..level..,  alpha = I(.3)),
                 size = 1, bins = 5, data = airport_df,
                 geom = "polygon") +
  geom_point(color="red", alpha = .2, size=airport_df$eigenvector*10) +
  world_theme +
geom_text_repel(data = subset(airport_df, eigenvector >= thres), aes(x=Longtitude, y= Latitude, la
bel=Name), color = "black", fontface = "italic", size = 2, max.overlaps = Inf) +
  ggtitle("By Eigenvector")

eigen_plot
```

By Eigenvector




```
for (i in top20_eigen_df$Name){
  if (i %nin% top20_degree_df$Name){
    print(i)
  }
}
```

```
## [1] "San Francisco International Airport"
## [1] "Lester B. Pearson International Airport"
## [1] "Narita International Airport"
## [1] "Leonardo da Vinci–Fiumicino Airport"
## [1] "Adolfo Suárez Madrid–Barajas Airport"
## [1] "Philadelphia International Airport"
```

3. Comunnity detection

We are using the quicker method fastgreedy, so we will have to remove direction from our graph.

3.1 Sizes of Communities

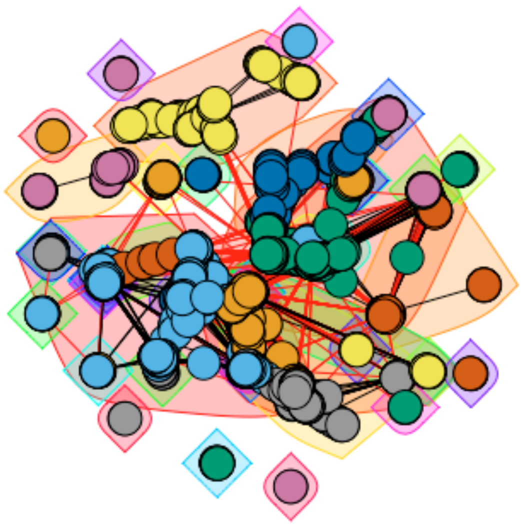
```
graph <- as.undirected(g)
graph <- simplify(graph)
fastgreedy_communities <- fastgreedy.community(graph)
V(graph)$community <- fastgreedy_communities$membership

sizes(fastgreedy_communities)
```

```
## Community sizes
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
## 732 898 800 183  74 155  52 178  37  12  17  12  19   7  25  15   7  18   9  13
##  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
##   8   6   6   6  22   5  10   8   4   4  23   4   7   3   3   7   3   4   4   3
##  41  42  43  44  45  46  47  48  49
##   3   4   3   2   2   2   2   2   2
```

We have obtained 49 communities, we will explore the biggest 4.

3.2 Initial Plotting

```
plot(fastgreedy_communities, graph, vertex.label = NA)
```



```
#Attaching community id to the airport_df
membership_vec <- membership(fastgreedy_communities)
membership_df <-as.data.frame(as.table(membership_vec))
names(membership_df)[1] <- "id"
names(membership_df)[2] <- "community id"
airport_df <-airport_df %>% left_join(membership_df, by = c("IATA" = "id"))
```

We will then take samples of 20 to see how are these communities formed.

```
sample_n(airport_df[airport_df$`community id` == 1,], 20)[,c("IATA", "Country","Name")]
```

```
##    IATA        Country                                                 Name
## 1   SLM          Spain                                    Salamanca Airport
## 2   MZH         Turkey                              Amasya Merzifon Airport
## 3   EGO         Russia                        Belgorod International Airport
## 4   LRS         Greece                                        Leros Airport
## 5   HER         Greece Heraklion International Nikos Kazantzakis Airport
## 6   OUA   Burkina Faso                                   Ouagadougou Airport
## 7   SFL     Cape Verde                                    São Filipe Airport
## 8   AKX     Kazakhstan                                       Aktobe Airport
## 9   JSY         Greece                                        Syros Airport
## 10  CEG United Kingdom                                     Hawarden Airport
## 11  AQJ         Jordan          Aqaba King Hussein International Airport
## 12  BIA         France                                Bastia-Poretta Airport
## 13  PSA          Italy                           Pisa International Airport
## 14  LDE         France                       Tarbes-Lourdes-Pyrénées Airport
## 15  PAS         Greece                               Paros National Airport
## 16  DTM        Germany                                     Dortmund Airport
## 17  LPL United Kingdom                         Liverpool John Lennon Airport
## 18  MQM         Turkey                                       Mardin Airport
## 19  ECN         Cyprus                          Ercan International Airport
## 20  NCE         France                            Nice-Côte d'Azur Airport
```

```
sample_n(airport_df[airport_df$`community id` == 2,], 20)[,c("IATA", "Country","Name")]
```

```
##    IATA      Country                                       Name
## 1   KET        Burma                           Kengtung Airport
## 2   YCU        China                  Yuncheng Guangong Airport
## 3   TIF Saudi Arabia                     Ta’if Regional Airport
## 4   LDH    Australia                   Lord Howe Island Airport
## 5   WXN        China                            Wanxian Airport
## 6   IXZ        India        Vir Savarkar International Airport
## 7   HDY     Thailand           Hat Yai International Airport
## 8   HET        China             Baita International Airport
## 9   HTN        China                              Hotan Airport
## 10  BJB         Iran                            Bojnord Airport
## 11  BDP        Nepal                          Bhadrapur Airport
## 12  COK        India            Cochin International Airport
## 13  KUU        India                       Kullu Manali Airport
## 14  WAE Saudi Arabia                   Wadi Al Dawasir Airport
## 15  HIN  South Korea                  Sacheon Air Base/Airport
## 16  MKY    Australia                             Mackay Airport
## 17  BDO    Indonesia Husein Sastranegara International Airport
## 18  DSN        China                    Ordos Ejin Horo Airport
## 19  TDX     Thailand                               Trat Airport
## 20  MWF      Vanuatu                       Maewo-Naone Airport
```

```
sample_n(airport_df[airport_df$`community id` == 3,], 20)[,c("IATA", "Country","Name")]
```

```
##    IATA                  Country
## 1   PLN            United States
## 2   PUQ                    Chile
## 3   BRL            United States
## 4   CHS            United States
## 5   XSC Turks and Caicos Islands
## 6   QBC                   Canada
## 7   DBQ            United States
## 8   UPN                   Mexico
## 9   YCD                   Canada
## 10  CAY            French Guiana
## 11  YGP                   Canada
## 12  YQL                   Canada
## 13  SHD            United States
## 14  TWF            United States
## 15  MHH                  Bahamas
## 16  CUE                  Ecuador
## 17  HOB            United States
## 18  LGA            United States
## 19  LAS            United States
## 20  CRP            United States
##                                                Name
## 1  Pellston Regional Airport of Emmet County Airport
## 2              Pdte. Carlos Ibañez del Campo Airport
## 3                    Southeast Iowa Regional Airport
## 4    Charleston Air Force Base-International Airport
## 5                              South Caicos Airport
## 6                               Bella Coola Airport
## 7                          Dubuque Regional Airport
## 8   Licenciado y General Ignacio Lopez Rayon Airport
## 9                                   Nanaimo Airport
## 10                       Cayenne-Rochambeau Airport
## 11                   Gaspé (Michel-Pouliot) Airport
## 12                        Lethbridge County Airport
## 13                Shenandoah Valley Regional Airport
## 14        Joslin Field Magic Valley Regional Airport
## 15          Leonard M Thompson International Airport
## 16                            Mariscal Lamar Airport
## 17                       Lea County Regional Airport
## 18                                La Guardia Airport
## 19                     McCarran International Airport
## 20              Corpus Christi International Airport
```

```
sample_n(airport_df[airport_df$`community id` == 4,], 20)[,c("IATA", "Country","Name")]
```

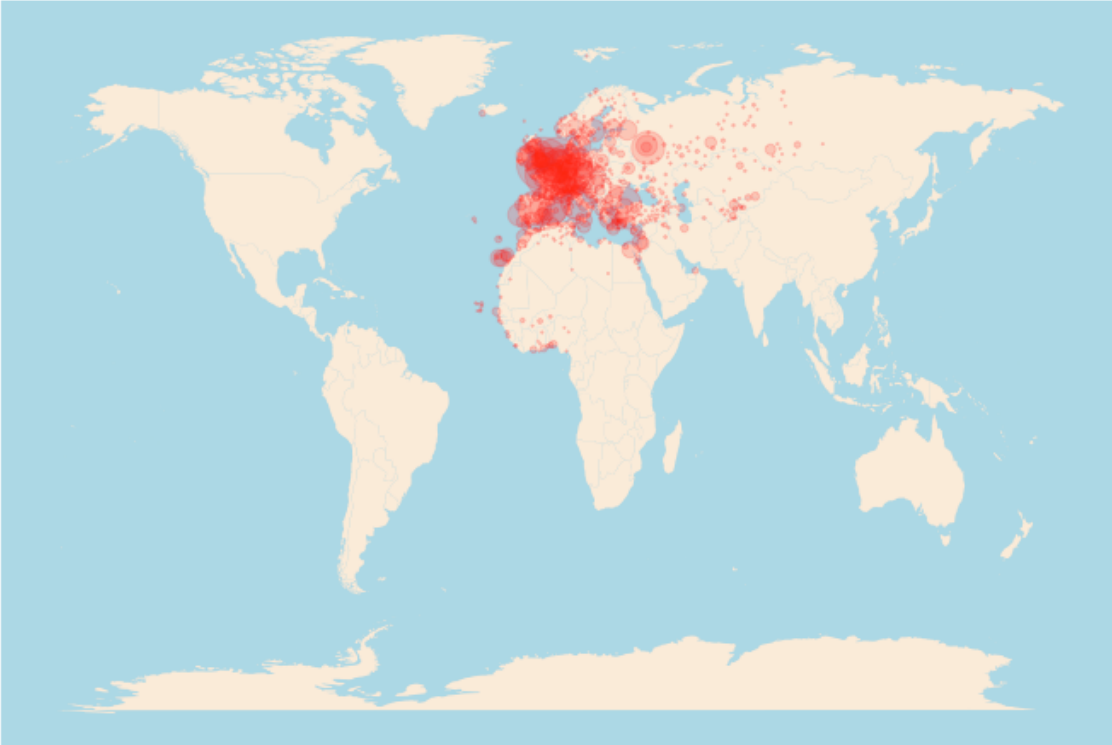
```
##    IATA       Country                       Name
## 1   HUS United States             Hughes Airport
## 2   SKK United States         Shaktoolik Airport
## 3   NLG United States      Nelson Lagoon Airport
## 4   KPN United States             Kipnuk Airport
## 5   BTI United States Barter Island LRRS Airport
## 6   SNP United States     St Paul Island Airport
## 7   ADK United States               Adak Airport
## 8   ADQ United States             Kodiak Airport
## 9   BKC United States           Buckland Airport
## 10  FYU United States         Fort Yukon Airport
## 11  AKN United States        King Salmon Airport
## 12  KAL United States             Kaltag Airport
## 13  HNS United States             Haines Airport
## 14  NUI United States            Nuiqsut Airport
## 15  SDP United States         Sand Point Airport
## 16  CYF United States          Chefornak Airport
## 17  RSH United States    Russian Mission Airport
## 18  GST United States           Gustavus Airport
## 19  DLG United States         Dillingham Airport
## 20  SMK United States         St Michael Airport
```

```
community_1 <- airport_df %>% dplyr::filter(`community id` == 1)
community_2 <- airport_df %>% dplyr::filter(`community id` == 2)
community_3 <- airport_df %>% dplyr::filter(`community id` == 3)
community_4 <- airport_df %>% dplyr::filter(`community id` == 4)
```

3.3 Graphing the Communities

```
community_1_plot <- ggplot(community_1, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  world_theme +
  geom_point(color="red", alpha = .2, size=community_1$degree/100) +
  ggtitle("Community 1")

community_1_plot
```

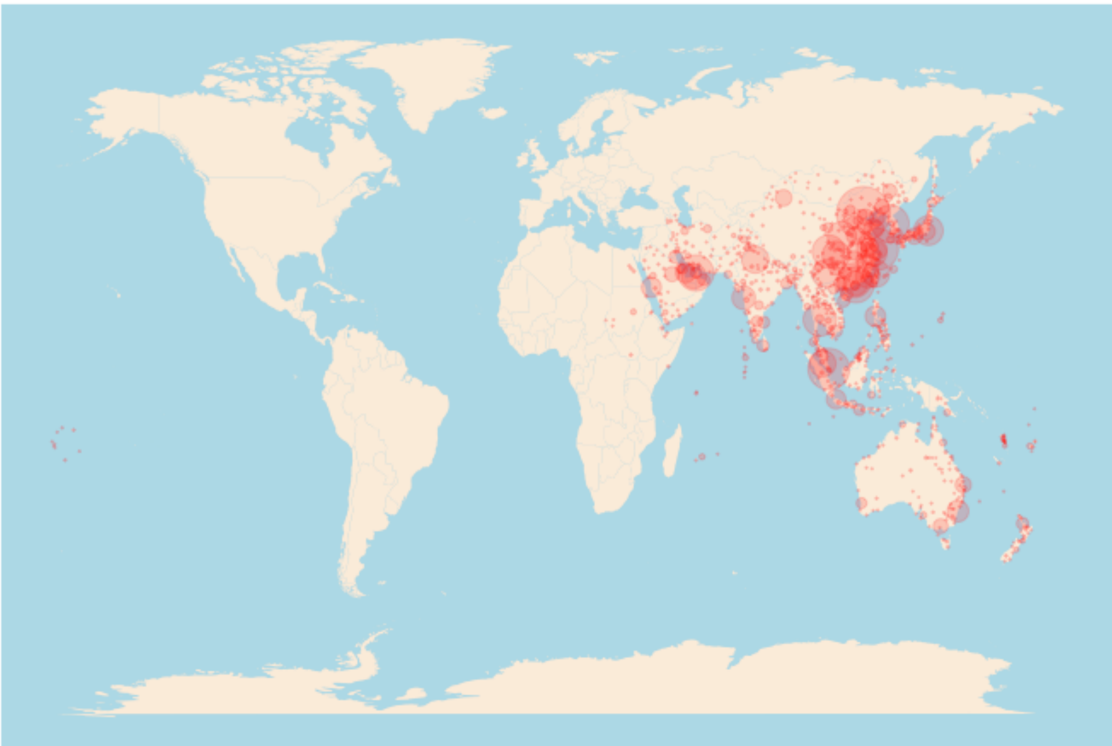
Community 1



Community 1 is focused on Europe, a bit of Middle East and some coastal part of Africa.

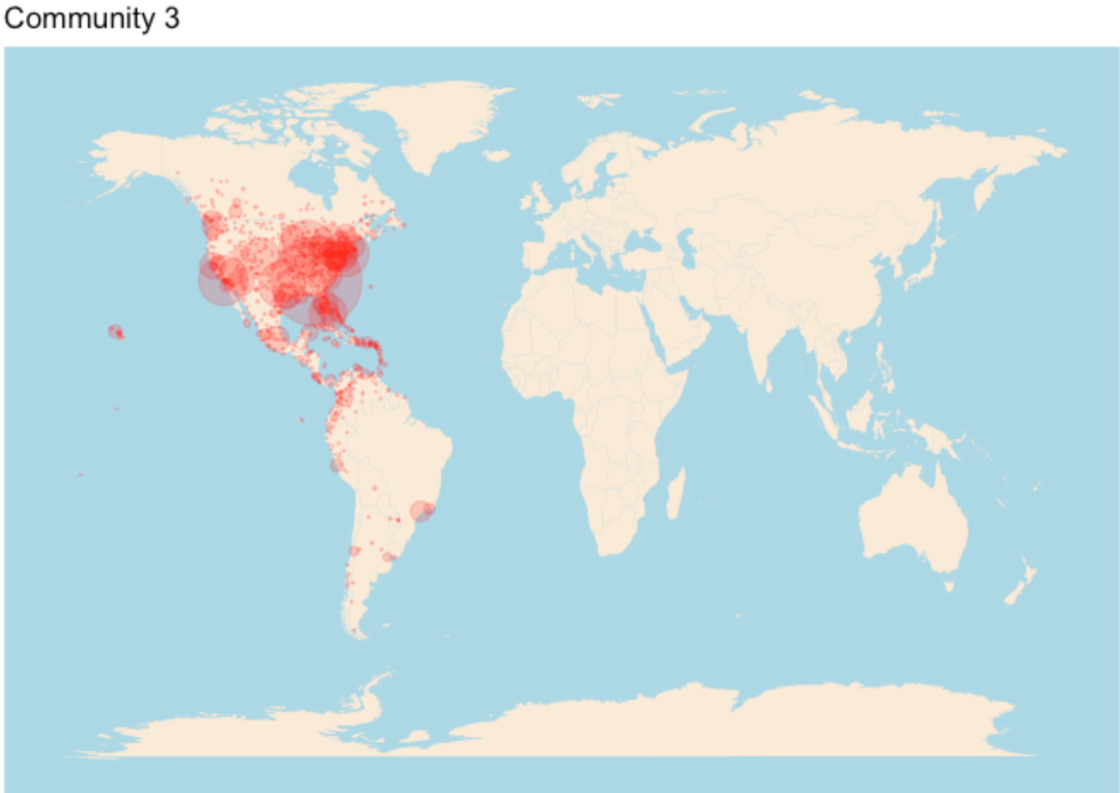
```
community_2_plot <- ggplot(community_2, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  world_theme +
  geom_point(color="red", alpha = .2, size=community_2$degree/100) +
  ggtitle("Community 2")

community_2_plot
```

Community 2



Community 2 is focused on the Asia-Pacific, Central Asia and Middle East.

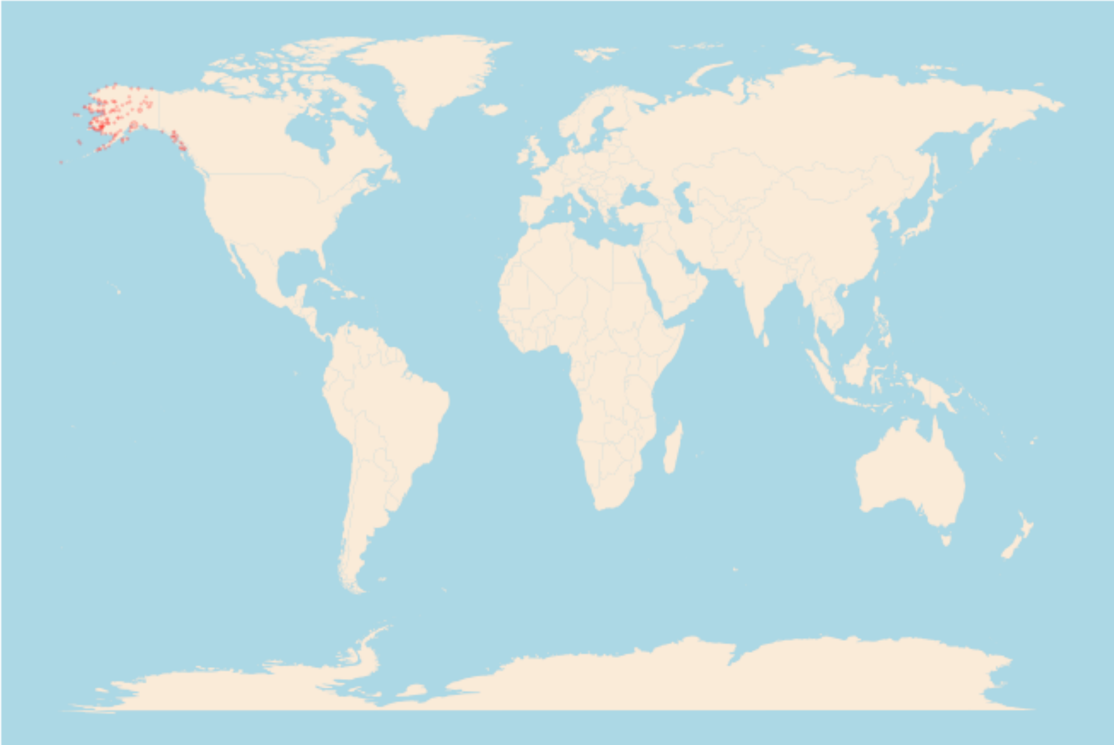
```
community_3_plot <- ggplot(community_3, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  world_theme +
  geom_point(color="red", alpha = .2, size=community_3$degree/100) +
  ggtitle("Community 3")

community_3_plot
```



Community 3 is focused in US and some parts of Central & South America.

```
community_4_plot <- ggplot(community_4, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  world_theme +
  geom_point(color="red", alpha = .2, size=community_4$degree/100) +
  ggtitle("Community 4")

community_4_plot
```

Community 4



Community 4 is mostly centralized in Alaska, with few routes.

4. Insights & Analysis

4.1 Where are these routes flying from and to?

```
country_origin_df <- airport_df %>% dplyr::select("IATA","Country") %>% dplyr::rename(source.airpo
rt=IATA)
df_1 <- merge(x = routes_df, y = country_origin_df, by = "source.airport", all.x = TRUE)
df_1 <- df_1 %>% dplyr::rename(Country_origin=Country)

country_destination_df <- airport_df %>% dplyr::select("IATA","Country") %>% dplyr::rename(destina
tion.airport=IATA)
df_2 <- merge(x = df_1, y = country_destination_df, by = "destination.airport", all.x = TRUE)
df_2 <- df_2 %>% dplyr::rename(Country_destination=Country)

df3 <- df_2 %>% dplyr::count(Country_origin, Country_destination, sort=TRUE)
df3 <- df3 %>%dplyr::rename(number_of_routes=n, source=Country_origin, target=Country_destination)
df3[1:30,]
```

```
##            source         target number_of_routes
## 1   United States  United States            10518
## 2           China          China             6877
## 3          Brazil         Brazil             1195
## 4          Canada         Canada             1167
## 5           India          India             1057
## 6          Russia         Russia              964
## 7       Australia      Australia              776
## 8           Japan          Japan              623
## 9       Indonesia      Indonesia              611
## 10          Spain          Spain              579
## 11         Mexico         Mexico              577
## 12 United Kingdom          Spain              518
## 13          Spain United Kingdom              512
## 14         France         France              483
## 15          Italy          Italy              425
## 16         Mexico  United States              373
## 17  United States         Mexico              369
## 18  United States         Canada              364
## 19         Canada  United States              363
## 20        Germany          Spain              354
## 21          Spain        Germany              353
## 22 United Kingdom United Kingdom              309
## 23         Turkey         Turkey              306
## 24           Iran           Iran              304
## 25         Norway         Norway              302
## 26       Malaysia       Malaysia              256
## 27    Philippines    Philippines              240
## 28         Greece         Greece              235
## 29       Colombia       Colombia              233
## 30        Germany          Italy              221
```

We could see most popular routes are domestic,and are from countries that are either big geographically or population-wise. Then some of the most popular internal routes are:

- UK to Spain
- Spain to UK
- Mexico to US

4.2 Diameter

Diameter: Which is the longest route?

```
diameter_routes <- diameter(g, directed = TRUE)
print(paste("The diameter of the route graph is", diameter_routes, ", which means one person can g
o to", diameter_routes, "cities in one go without repeating the places this person has been."))
```

```
## [1] "The diameter of the route graph is 14 , which means one person can go to 14 cities in one
go without repeating the places this person has been."
```

```
diameter_stops <- get_diameter(g)
diameter_stops <- as.vector(names(diameter_stops))
diameter_df <- airport_df[match(diameter_stops, airport_df$IATA),]
diameter_df <- diameter_df[complete.cases(diameter_df),]
diameter_df
```

```
##      Airport.ID                                               Name
## 2340       5535                                    Salluit Airport
## 2318       5504                                   Ivujivik Airport
## 2319       5506                                   Akulivik Airport
## 2879       6727                                 Puvirnituq Airport
## 40           62                          La Grande Rivière Airport
## 90          146 Montreal / Pierre Elliott Trudeau International Airport
## 759        1665                  Geneva Cointrin International Airport
## 298         609                           Copenhagen Kastrup Airport
## 9             9                                Kangerlussuaq Airport
## 7             7                                 Narsarsuaq Airport
## 2275       5442                                  Qaqortoq Heliport
## 2277       5444                                Nanortalik Heliport
##                  City     Country IATA Latitude Longtitude degree in_degree
## 2340          Salluit      Canada  YZG 62.17940  -75.66720      4         2
## 2318         Ivujivik      Canada  YIK 62.41730  -77.92530      4         2
## 2319         Akulivik      Canada  AKV 60.81860  -78.14860      4         2
## 2879       Puvirnituq      Canada  YPX 60.05060  -77.28690      8         4
## 40   La Grande Riviere      Canada  YGL 53.62530  -77.70420      6         3
## 90           Montreal      Canada  YUL 45.47060  -73.74080    371       186
## 759            Geneva Switzerland  GVA 46.23810    6.10895    329       163
## 298        Copenhagen     Denmark  CPH 55.61790   12.65600    457       228
## 9          Sondrestrom   Greenland  SFJ 67.01222  -50.71160     16         8
## 7          Narssarssuaq   Greenland  UAK 61.16050  -45.42600     10         5
## 2275         Qaqortoq   Greenland  JJU 60.71568  -46.02992     14         7
## 2277       Nanortalik   Greenland  JNN 60.14188  -45.23298      8         4
##      out_degree  betweenness    closeness  eigenvector degree_diff community id
## 2340          2 9.583333e+00 5.330803e-06 5.153692e-12           0            5
## 2318          2 4.489352e+03 5.428292e-06 6.067198e-10           0            5
## 2319          2 1.122435e+04 5.529444e-06 1.072068e-07           0            5
## 2879          4 1.797143e+04 5.634438e-06 1.894363e-05           0            5
## 40            3 1.220500e+04 5.741847e-06 1.665524e-03           0            5
## 90          185 3.114321e+05 5.854869e-06 2.926011e-01          -1            3
## 759         166 1.250666e+04 5.850862e-06 1.878552e-01           3            1
## 298         229 3.178263e+05 5.868028e-06 2.566127e-01           1            1
## 9             8 2.369665e+05 5.757417e-06 1.454688e-03           0            7
## 7             5 7.741508e+04 5.654669e-06 1.040234e-05           0            7
## 2275          7 6.062100e+04 5.549359e-06 5.921119e-08           0            7
## 2277          4 2.023300e+04 5.447543e-06 3.388999e-10           0            7
```

```
diameter_plot <- ggplot(diameter_df, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  world_theme +
  geom_point(color="red", alpha = .5, size=2) +
  geom_text_repel(aes(label=Name), color = "black", fontface = "italic", size = 2, max.overlaps =
Inf)
```

```
x1 <- diameter_df[1,"Longtitude"]
x2 <- diameter_df[2,"Longtitude"]
x3 <- diameter_df[3,"Longtitude"]
x4 <- diameter_df[4,"Longtitude"]
x5 <- diameter_df[5,"Longtitude"]
x6 <- diameter_df[6,"Longtitude"]
x7 <- diameter_df[7,"Longtitude"]
x8 <- diameter_df[8,"Longtitude"]
x9 <- diameter_df[9,"Longtitude"]

y1 <- diameter_df[1,"Latitude"]
y2 <- diameter_df[2,"Latitude"]
y3 <- diameter_df[3,"Latitude"]
y4 <- diameter_df[4,"Latitude"]
y5 <- diameter_df[5,"Latitude"]
y6 <- diameter_df[6,"Latitude"]
y7 <- diameter_df[7,"Latitude"]
y8 <- diameter_df[8,"Latitude"]
y9 <- diameter_df[9,"Latitude"]


diameter_plot +
geom_curve(aes(x = x1, y = y1, xend = x2, yend = y2, colour = "black")) +
geom_curve(aes(x = x2, y = y2, xend = x3, yend = y3, colour = "black")) +
geom_curve(aes(x = x3, y = y3, xend = x4, yend = y4, colour = "black")) +
geom_curve(aes(x = x4, y = y4, xend = x5, yend = y5, colour = "black")) +
geom_curve(aes(x = x5, y = y5, xend = x6, yend = y6, colour = "black")) +
geom_curve(aes(x = x6, y = y6, xend = x7, yend = y7, colour = "black")) +
geom_curve(aes(x = x7, y = y7, xend = x8, yend = y8, colour = "black")) +
geom_curve(aes(x = x8, y = y8, xend = x9, yend = y9, colour = "black")) +
ggtitle("Diameter Path")
```

Diameter Path

4.3 Zoom in on Specific Cities

Where are the the places connected to Madrid

```
#CREATING SUBFRAMES WITH THE LATITUDE AND LONGITUDE FOR THE AIRPORT OF DEPARTURE AND THE AIRPORT O
F ARRIVAL

coords_origin <- airport_df %>% dplyr::select('Airport.ID', 'Latitude', 'Longtitude') %>% dplyr::r
ename(SourceLat=Latitude,SourceLong=Longtitude)

coords_destiny <- airport_df %>% dplyr::select('Airport.ID', 'Latitude', 'Longtitude') %>% dplyr::
rename(DestLat=Latitude, DestLong=Longtitude)

flights_to_from <- routes_df %>%
  filter((routes_df$source.airport=="MAD") |     (routes_df$destination.airport=="MAD"))

#MERGING WITH ROUTES DATAFRAME BUT USING THE AIRPORT OF DEPARTURE AS COMMON COLUMN
flights_coords_origin <- merge(flights_to_from, coords_origin, by.x='source.airport.id', by.y='Air
port.ID' )

#MERGING THE DATAFRAME FLIGHTS_COORDS_ORIGIN WITH THE COORDINATES OF ARRIVAL USING THE AIRPORT OF
ARRIVAL AS COMMON COLUMN
flights_coords_destination <- merge(flights_to_from, coords_destiny, by.x='destination.airport.id
', by.y='Airport.ID')

flights_with_coords <- merge(flights_coords_origin, flights_coords_destination)

#SUBFRAMING AND EXTRACTING ONLY COORDINATES OF ORIGIN AND DESTINATION
coords <- flights_with_coords %>% dplyr::select('SourceLat','SourceLong','DestLat','DestLong')
```

```
#CREATING DATAFRAME WITH THE COORDINATES OF ORIGIN
source_df<-data.frame(SourceLong=coords$SourceLong,SourceLat=coords$SourceLat)
#TRANSFORMING THEM INTO SPATIALPOINTS
source_sp<-SpatialPoints(source_df, proj4string=CRS("+proj=longlat"))
#CREATING A DATAFRAME OUT OF THOSE POINTS
source_spdf <- SpatialPointsDataFrame(source_sp, data = source_df)
```

```
#CREATING DATAFRAME WITH THE COORDINATES OF DESTINATION
dest_df<-data.frame(SourceLong=coords$DestLong,SourceLat=coords$DestLat)
#TRANSFORMING THEM INTO SPATIALPOINTS
dest_sp<-SpatialPoints(dest_df, proj4string=CRS("+proj=longlat"))
#CREATING A DATAFRAME OUT OF THOSE POINTS
dest_spdf <- SpatialPointsDataFrame(dest_sp, data = dest_df)
```

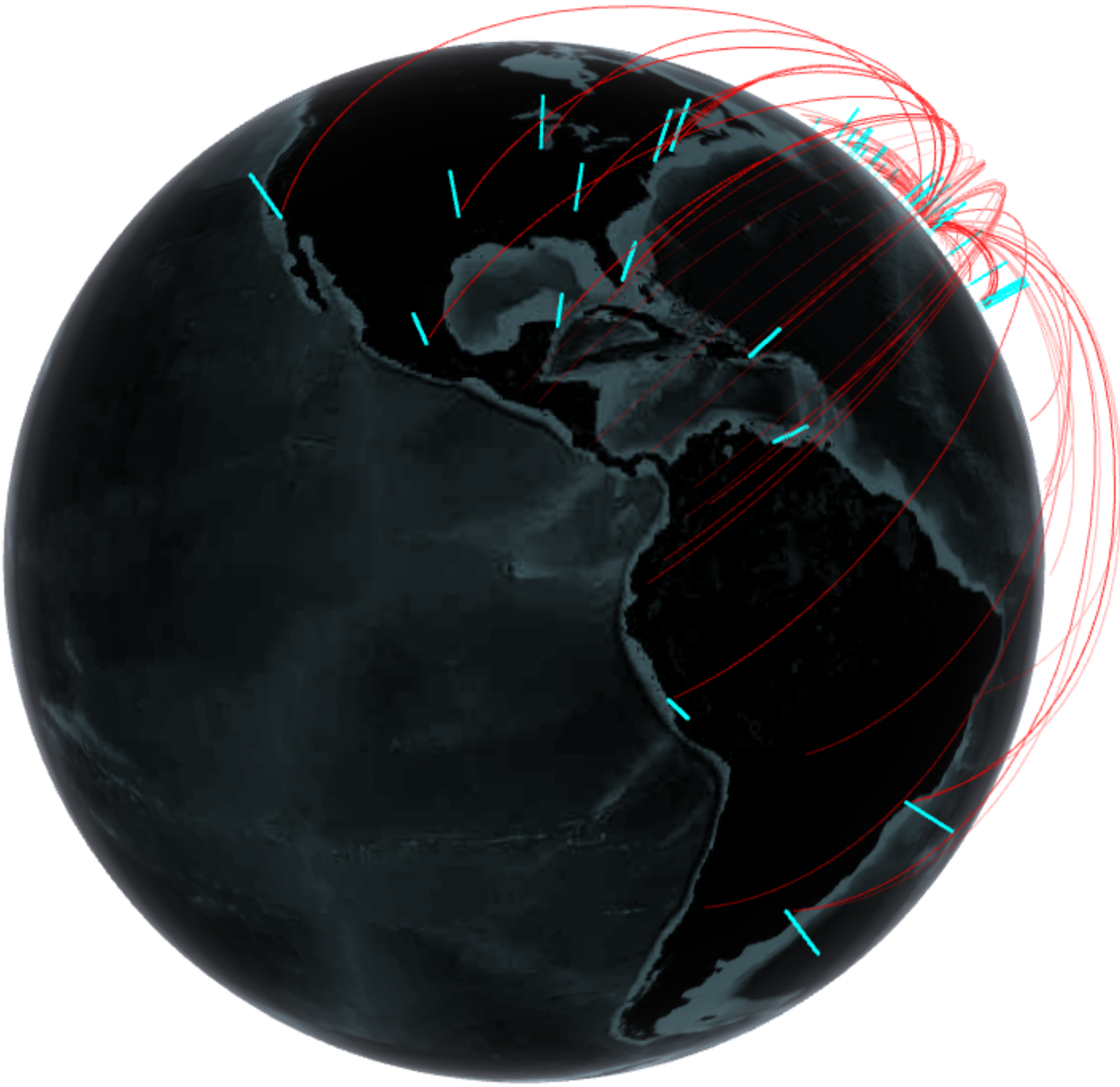
```
comb_df<-data.frame(coords)
comb_df$distance<-distHaversine(source_sp,dest_sp)
```

```
#ALLOWS US TO RETRIEVE THE MOST COMMON FLIGHTS BY COORDINATES
source_da   <- factor(sprintf("%.2f:%.2f",comb_df[,2], comb_df[,1]))
freq <- sort(table(source_da), decreasing=TRUE)
frequent_destinations <- names(freq)[1:50]
idx <- source_da %in% frequent_destinations
LongLat <- unique(comb_df[idx,1:2])
frequent_flights <-comb_df[idx,]
```

```
#PLOTTING THE DATAFRAME SO WE GET THE GLOBE
(earth <- system.file("images/world.jpg",  package="threejs"))
```

```
## [1] "/Library/Frameworks/R.framework/Versions/4.0/Resources/library/threejs/images/world.jpg"
```

```
test_df <- data.frame(origin_lat = comb_df[,1], origin_long = comb_df[,2], dest_lat = comb_df[,3],
dest_long = comb_df[,4])
#
globejs(img=earth, lat=LongLat[,1], long=LongLat[,2], arcs=test_df,
        arcsHeight=0.3, arcsLwd=2, arcsColor="red", arcsOpacity=0.15,
        atmosphere=TRUE,bg="white", height = 800 , width = 800)
```



4.4 Travelling Between 2 Specific Stops

If I'm a consultant based in Madrid and I go to Shanghai every month, which airline should I pick?

```
distances(g, "MAD", "PVG")
```

```
##      PVG
## MAD    2
```

```
n_mad <- neighbors(g, "MAD", mode = c('in'))
n_pvg <- neighbors(g, "PVG", mode = c('out'))
middle_stops <- as.table(intersection(n_mad, n_pvg))
names(middle_stops)
```

```
##  [1] "ZRH" "ORD" "BKK" "MUC" "AMS" "ICN" "PEK" "LAX" "DXB" "JFK" "CDG" "FCO"
## [13] "FRA" "HEL" "LHR" "MXP" "SVO" "CPH" "IST" "DOH" "EWR"
```

```
tempdf <- routes_df %>% dplyr::select("airline","source.airport", "destination.airport") %>% dply
r::rename(source = source.airport, dest = destination.airport)

tempdf1 <- tempdf %>% dplyr::filter(source == "MAD") %>% dplyr::rename(airline1 = airline)

tempdf2 <- tempdf %>% dplyr::filter(dest == "PVG") %>% dplyr::rename(airline2 = airline)
```

```
sqldf("select tempdf1.*, tempdf2.* from tempdf1, tempdf2 where (tempdf1.dest = tempdf2.source) and
airline1 = airline2")
```

```
##    airline1 source dest airline2 source dest
## 1        AA    MAD  LAX       AA    LAX  PVG
## 2        AA    MAD  ORD       AA    ORD  PVG
## 3        AF    MAD  CDG       AF    CDG  PVG
## 4        AY    MAD  HEL       AY    HEL  PVG
## 5        AZ    MAD  FCO       AZ    FCO  PVG
## 6        BA    MAD  LHR       BA    LHR  PVG
## 7        CA    MAD  PEK       CA    PEK  PVG
## 8        DL    MAD  JFK       DL    JFK  PVG
## 9        EK    MAD  DXB       EK    DXB  PVG
## 10       KE    MAD  ICN       KE    ICN  PVG
## 11       KL    MAD  AMS       KL    AMS  PVG
## 12       LH    MAD  FRA       LH    FRA  PVG
## 13       LH    MAD  MUC       LH    MUC  PVG
## 14       LX    MAD  ZRH       LX    ZRH  PVG
## 15       MU    MAD  AMS       MU    AMS  PVG
## 16       QR    MAD  DOH       QR    DOH  PVG
## 17       SK    MAD  CPH       SK    CPH  PVG
## 18       SU    MAD  SVO       SU    SVO  PVG
## 19       TG    MAD  BKK       TG    BKK  PVG
## 20       TK    MAD  IST       TK    IST  PVG
## 21       UA    MAD  EWR       UA    EWR  PVG
```

From the list we could see that American Airline(AA) and Lufthansa Airline (LF) are the only two airlines that have more than 1 routes fully operated by themselves. As there could be uncertainty as airports, given more than 1 choice as the pit stop could be better options.

4.5 Adding Passenger Volume

Are the busiest airport really busy? We wanted to add the passenger volume to the data set to evaluate their relation with degree relationship. As the free & available data only has ranked 20 airports, we will do it on a small scale.

```
passenger_url <- "https://gist.githubusercontent.com/hannahbhchou/01cbc0081c8a080350e50d0ead1a1fcc
/raw/33f3a9b29ae6a7323ace128f94775025d23485cb/passenger_2017.csv"

passenger_df <- read.csv(passenger_url, header = TRUE)
```

```
passenger_df <- passenger_df %>% left_join(airport_df, by = c("IATA" = "IATA"))
passenger_df$v_d_ratio <- with(passenger_df, Volume / degree)
```

```
passenger_df[,c("IATA", "Name", "Volume", "degree", "v_d_ratio")]
```

```
##    IATA                                         Name    Volume degree
## 1   ATL Hartsfield Jackson Atlanta International Airport 103902992   1826
## 2   PEK           Beijing Capital International Airport  95786442   1069
## 3   DXB                    Dubai International Airport  88242099    710
## 4   HND              Tokyo Haneda International Airport  85408975    315
## 5   LAX              Los Angeles International Airport  84557968    990
## 6   ORD            Chicago O'Hare International Airport  79828183   1108
## 7   LHR                        London Heathrow Airport  78014598   1051
## 8   HKG                Hong Kong International Airport  72664075    710
## 9   PVG          Shanghai Pudong International Airport  70001237    825
## 10  CDG          Charles de Gaulle International Airport  69471442   1041
## 11  AMS                      Amsterdam Airport Schiphol  68515425    903
## 12  DFW          Dallas Fort Worth International Airport  67092194    936
## 13  CAN          Guangzhou Baiyun International Airport  65887473    674
## 14  FRA                       Frankfurt am Main Airport  64500386    990
## 15  IST                                Istanbul Airport  64119374    719
## 16  DEL             Indira Gandhi International Airport  63451503    527
## 17  CGK            Soekarno-Hatta International Airport  63015620    367
## 18  SIN                        Singapore Changi Airport  62220000    820
## 19  ICN                  Incheon International Airport  62157834    740
## 20  DEN                   Denver International Airport  61379396    735
##    v_d_ratio
## 1   56901.97
## 2   89603.78
## 3  124284.65
## 4  271139.60
## 5   85412.09
## 6   72047.10
## 7   74228.92
## 8  102343.77
## 9   84849.98
## 10  66735.29
## 11  75875.33
## 12  71679.69
## 13  97755.89
## 14  65151.91
## 15  89178.55
## 16 120401.33
## 17 171704.69
## 18  75878.05
## 19  83997.07
## 20  83509.38
```

One thing we've noticed that all of the top 20 passenger volume airports are scattered among the most popular communities, but maybe because of their geography they are separated.

```
passenger_plot <- ggplot(passenger_df, (aes(x = Longtitude, y= Latitude))) +
  borders("world", colour=NA, fill="antiquewhite")  +
  world_theme +
  geom_point(color="red", alpha = .2, size=passenger_df$v_d_ratio/18000) +
  geom_text_repel( aes(x=Longtitude, y= Latitude, label=Name), color = "black", fontface = "itali
c", size = 2, max.overlaps = Inf) +
  ggtitle("Top 20 Passenger Volume Airports")

passenger_plot
```

We could see that though Tokyo Haneda airport and Soekarno-Hatta International Airport are the highest in terms of volume/degree ratio, which means for every route they serve more passengers.