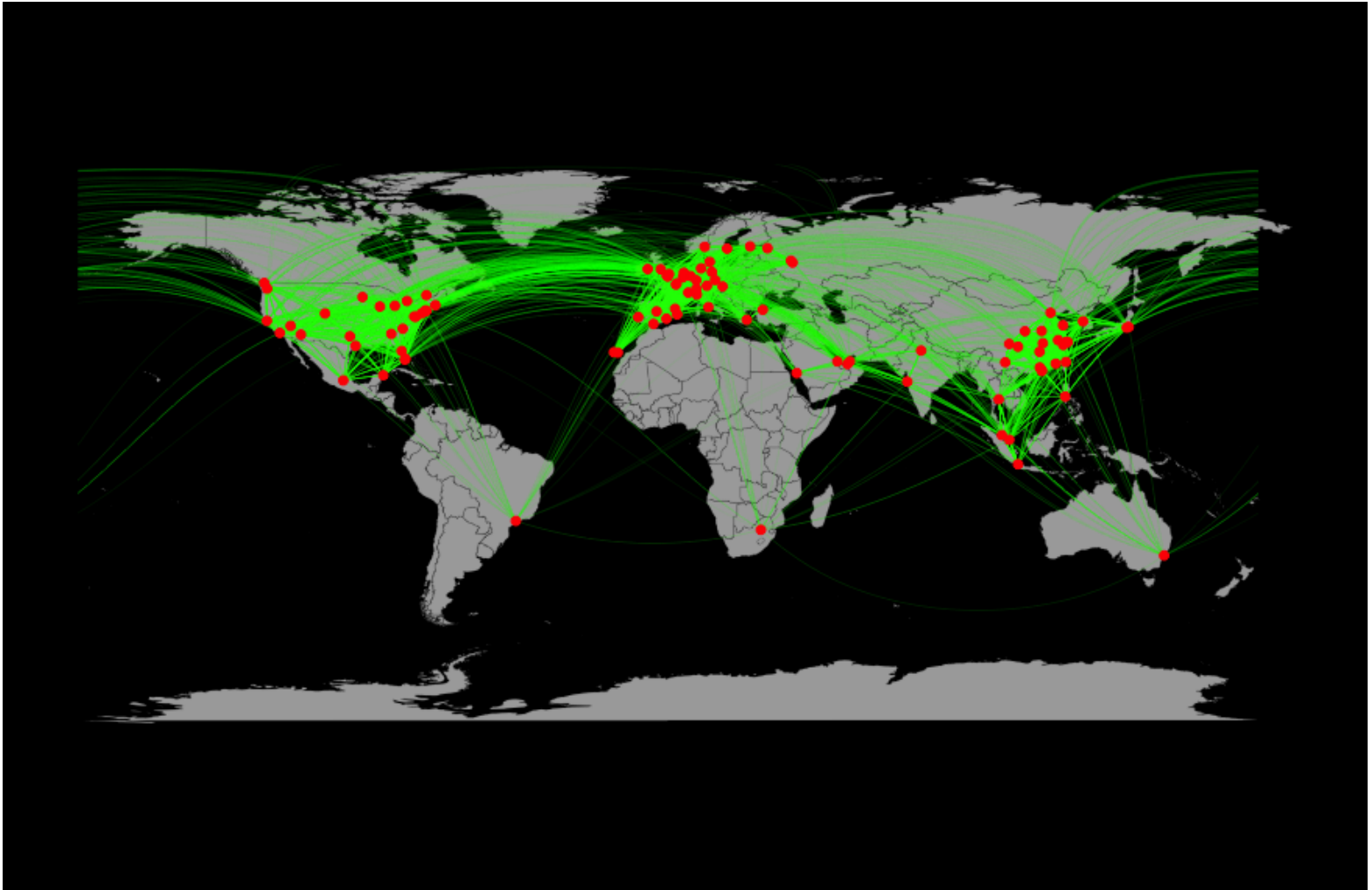


Homework 5

Yiwei He (yh9vhg), Da Lin (dl2de), Ziyue Jin (zj5qj)

Problem 1



Appendix

cleaning

```
airports <- read.csv("~/Desktop/Homework 5/airport/airports.txt",header=F)
colnames(airports) <- c("AirportID","Name","City","Country","IATA","ICAO",
  'Latitude','Longitude','Altitude','Timezone','DST',
  "Tz Timezone")
```

```
airlines <- read.csv("~/Desktop/Homework 5/airport/airlines.dat",header=F)
colnames(airlines) <- c("Airline ID","Name","Alias","IATA","ICAO","Callsign",
  "Country","Active")
```

```
routes <- read.csv("~/Desktop/Homework 5/airport/routes.dat",header=F)
colnames(routes) <- c("Airline","Airline ID","Source Airport",
  "Source Airport ID","Destination Airport",
  "Destination Airport ID","Codeshare","Stops","Equipment")
```

```
source_count <- as.data.frame(table(routes$`Source Airport`))
dest_count <- as.data.frame(table(routes$`Destination Airport`))
colnames(source_count) <- c("Source Airport","Count")
colnames(dest_count) <- c("Destination Airport","Count")
```

```
top_100 <- merge(source_count,dest_count,by.x="Source Airport",by.y="Destination Airport")
top_100$Count <- top_100$Count.x + top_100$Count.y
top_100 <- as.vector(top_100[order(-top_100$Count),])[1:100,c(1,4)]
head(top_100)
```

top 100 airports with largest flights

```
info <- unique(merge(top_100,routes[,3:4],by="Source Airport"))
ordered_ID <- info[order(-info$Count),]
dim(ordered_ID)
head(ordered_ID)
```

airport coordinates

```
airport.index <- which(
  unlist(lapply(airports$AirportID,function(x){
    any(ordered_ID$`Source Airport ID`==x)
  })))
)
airport_info <- airports[airport.index,]
airport_info
dim(airport_info)
```

flight source coordinates

```
flights.source.index <- which(
  unlist(lapply(routes$`Source Airport ID`,function(x){
```

```

    any(ordered_ID$`Source Airport ID`==x)
  )))
)

#### flight target coordinates
flights.target.index <- which(
  unlist(lapply(routes$`Destination Airport ID`,function(x){
    any(ordered_ID$`Source Airport ID`==x)
  })))
)

#### flights info
flights <- routes[intersect(flights.source.index,flights.target.index),]
dim(flights)
head(flights)

source_info <- routes[flights.source.index,]
target_info <- routes[flights.target.index,]
head(source_info)
head(target_info)

## top-100 to top-100
node <- merge(flights,airports,by.x="Source Airport ID",by.y="AirportID",all.x=TRUE)
node <- node[,c(1,4,5,6,15,16)]
names(node)[5:6] <- c("Source Latitude","Source Longitude")
node <- merge(node,airports,by.x="Destination Airport ID",by.y="AirportID",all.x=TRUE)
node <- node[,c(1,2,3,4,5,6,12,13)]
names(node)[7:8] <- c("Target Latitude","Target Longitude")
dim(node)
head(node)

#### plot setup
library(maps)
library(RColorBrewer)
library("geosphere")

cols <- brewer.pal(8,"Purples")
col1 <- adjustcolor("Green",alpha=0.3)

edge.col <- colorRampPalette(c(col1,col2),alpha=TRUE)
edge.colors <- edge.col(100)

pdf("~/Desktop/Hw5.pdf")
map(database="world",col="dark gray",fill=TRUE,bg="black",lwd=0.1)

```

```

for (i in 1:nrow(airport_info)) {
  points(x=airport_info$Longitude[i],y=airport_info$Latitude[i],col="red",pch=19,cex=0.8)
}

for( i in 1:nrow(node)){
  arc <- gclIntermediate(c(node$`Source Longitude`[i],node$`Source Latitude`[i]),
                        c(node$`Target Longitude`[i],node$`Target Latitude`[i]),
                        n=1000,addStartEnd=TRUE,breakAtDateLine=TRUE)
  if (is.list(arc)) {
    arc1<-arc[[1]]
    arc2<-arc[[2]]
    lines(arc1,col=col1,lwd=0.1)
    lines(arc2,col=col1,lwd=0.1)
  }
  else {
    lines(arc,col=col1,lwd=0.1)
  }
}

dev.off()

```