

# Russian Passenger Lists

*Daniel Dowler*

*Thursday, March 17, 2016*

The National Archives provides access to passenger lists for ships leaving from Russia, Ireland, Germany, and Italy, between 1834 and ca. 1900 (see <https://aad.archives.gov/aad/series-list.jsp?cat=GP44>). The data files are accompanied by header files with ship information, as well as code files, which help with interpreting the data.

## Header Files

The header files contain information about each ship's journey. The manifest identification number is unique for each journey, and it is the foreign key for linking with the passenger lists tables. The header file for the Russian, German, and Italian data is the same. Note that some ships either made many journeys, or there were many ships with the same names (i.e. America: 230). We may be able to distinguish some ships, for example, if they make the same trip over and over.

Some ports were very busy for departures—port 4 (Liverpool) had 8,118 listed departures. On 5/4/1849, there were 30 ships from Russian ports, which arrived in the US.

```
hdrs <- read.csv("Data/manifest_headers.csv",header=T)
head(hdrs)
```

```
## Manifest.Identification.Number Ship.Name
## 1 1 MASSILIA
## 2 2 LOUISA
## 3 4 ITALIA
## 4 5 UNITED STATES
## 5 6 TEUTONIC
## 6 7 KENT
## Port.of.Departure..Embarkation. Date.of.Arrival
## 1 81 4/4/1894 0:00:00
## 2 4 5/30/1848 0:00:00
## 3 1 4/7/1894 0:00:00
## 4 4 5/31/1848 0:00:00
## 5 80 4/15/1897 0:00:00
## 6 4 5/31/1848 0:00:00
```

```
summary(hdrs)
```

```
## Manifest.Identification.Number Ship.Name
## Min. : 1 AMERICA : 230
## 1st Qu.: 16456 NEW YORK : 217
## Median : 36775 HERMANN : 213
## Mean : 247019 COLUMBIA : 204
## 3rd Qu.: 60252 WISCONSIN: 200
## Max. : 9999999 WYOMING : 191
## (Other) : 38216
## Port.of.Departure..Embarkation. Date.of.Arrival
```

```
## 4      : 8118      5/4/1849 0:00:00 : 30
## 14     : 5503      5/3/1849 0:00:00 : 29
## 80     : 3691      5/18/1850 0:00:00: 26
## 16     : 3499      5/24/1847 0:00:00: 23
## 23     : 2398      5/29/1848 0:00:00: 23
## 22     : 1922      5/2/1849 0:00:00 : 22
## (Other):14340      (Other)          :39318
```

The ports above have numerical codes. A separate table shows their nominal values.

```
port <- read.csv("russian_codes/port_cl_4549.csv",header=F)
head(port)
```

```
## V1      V2 V3
## 1 1      NAPLES NA
## 2 2  MARSEILLES NA
## 3 3 LONDONDERRY NA
## 4 4  LIVERPOOL NA
## 5 5  QUEENSTOWN NA
## 6 6  GLASGOW NA
```

```
dim(port)
```

```
## [1] 432 3
```

## Russian Passenger Lists

```
rus_pass_list <- read.csv("Data/Russian_passenger_data.csv",header=T)
head(rus_pass_list)
```

```
## Last.Name First.Name Age Sex Occupation Literacy Country.of.Origin
## 1 ABRAMOWITZ A. 22 M TLR U 44
## 2 ARMANOWITZ VINCENT 33 M LABR U 44
## 3 BATZ BENGAN 41 M TLR U 44
## 4 BEER HIRSCH 18 M LABR U 44
## 5 BERGMAN JANKEL 32 M GZR U 44
## 6 BERLINER SROEL 17 M SHMK U 44
## City.Town.of.Last.Residence Destination.City.Country
## 1 U NEW YORK
## 2 U NEW YORK
## 3 U PHILADELPHIA
## 4 U NEW YORK
## 5 U NEW YORK
## 6 U NEW YORK
## Transit.and.or.Travel.Compartment Manifest.Identification.Number
## 1 S13 294
## 2 S13 294
## 3 S13 294
## 4 S13 294
## 5 S13 294
## 6 S13 294
```

```
summary(rus_pass_list)
```

```
##      Last.Name      First.Name      Age      Sex
## LEWIN      : 2353 JOSEF      : 10568 Min.      : 1.00 F :199961
## U          : 1960 JACOB      : 8823 1st Qu.: 16.00 FM:      11
## KAPLAN      : 1573 MARIA      : 7672 Median   : 23.00 M :326717
## GOLDBERG: 1571 MOSES      : 7280 Mean      : 68.27 MF:       2
## KATZ        : 1478 ANNE      : 6765 3rd Qu.: 32.00 P :       1
## SCHMIDT     : 1394 JAN       : 6616 Max.      :911.00 U :      702
## (Other) :517065 (Other):479670
##      Occupation      Literacy      Country.of.Origin City.Town.of.Last.Residence
## LABR      :153488 N: 27977 Min.      : 10 U      :373269
## CH        :102553 U:470788 1st Qu.: 44 HAMBURG: 13909
## U          : 48418 Y: 28629 Median   : 44 KOWNO   : 7824
## W          : 41992 Mean      : 40 SUWALKI: 6140
## FMR        : 35745 3rd Qu.: 44 RUSSIA  : 5759
## INF        : 23238 Max.      :119 WARSAW  : 4518
## (Other):121960 (Other):115975
## Destination.City.Country Transit.and.or.Travel.Compartment
## USA          :228121 S13      :503791
## NEW YORK     :177029 S09      : 8988
## PA           : 15877 T13      : 3400
## PHILADELPHIA: 9922 T09      : 3023
## BOSTON       : 8964 I13      : 3011
## BALTIMORE    : 6992 C09      : 2387
## (Other)      : 80489 (Other): 2794
## Manifest.Identification.Number
## Min.      : 294
## 1st Qu.: 38713
## Median : 41864
## Mean      : 44919
## 3rd Qu.: 43758
## Max.      :903276
##
```

## Variables Detail

### Age

Let's take a look at the age codes. Column V1 is the code, column V2 describes the age, and column V3 is empty. The age codes correspond to actual ages (i.e. 2 is 2 years old), except for codes over 99; these codes cover infants, and unknown ages.

```
setwd("E:/w209_Final_Project_Immigration/")
rus_age <- read.csv("russian_codes/rus_age_cl_4629.csv",header=F)
head(rus_age)
```

```
##      V1      V2 V3
## 1  1 age 01
## 2  2 age 02
```

```
## 3 3 age 03
## 4 4 age 04
## 5 5 age 05
## 6 6 age 06
```

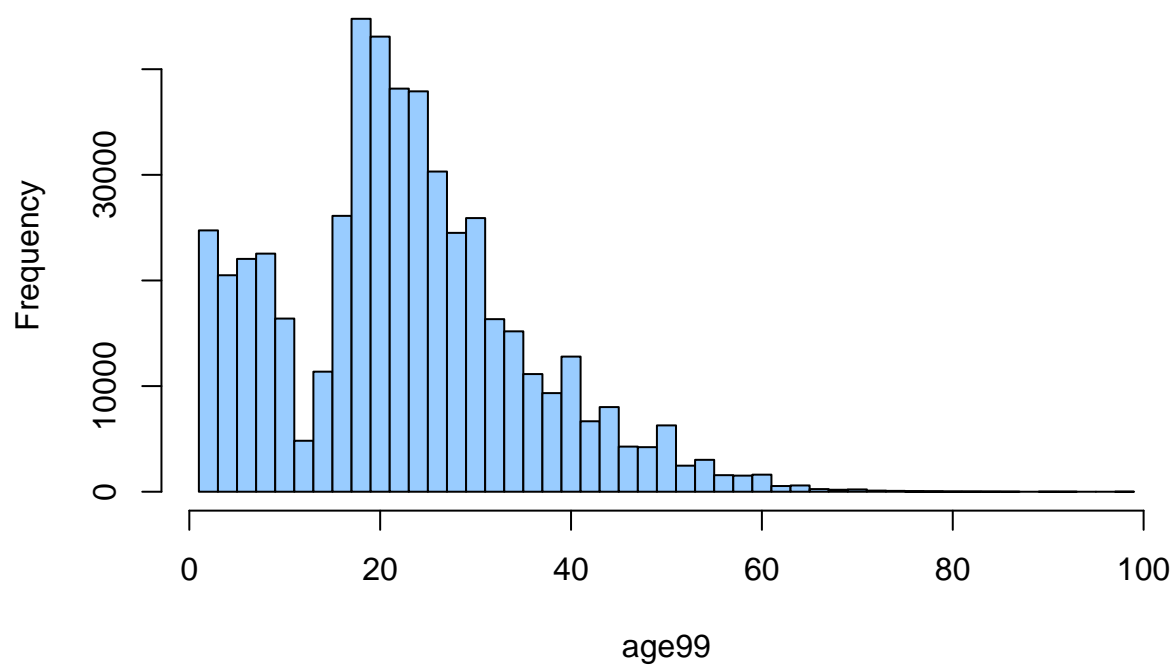
```
infant_n_unknown <- rus_age[rus_age$V1 >= 100,c(1,2)]
infant_n_unknown
```

```
##      V1      V2
## 154 800      Unknown
## 166 900 Infant in months: 00
## 167 901 Infant in months: 01
## 168 902 Infant in months: 02
## 169 903 Infant in months: 03
## 170 904 Infant in months: 04
## 171 905 Infant in months: 05
## 172 906 Infant in months: 06
## 173 907 Infant in months: 07
## 174 908 Infant in months: 08
## 175 909 Infant in months: 09
## 177 910 Infant in months: 10
## 178 911 Infant in months: 11
```

The codes for infant ages (i.e. 2 months) are in the 900 range. The 800 code stands for unknown. All other codes are equal to the ages themselves. Most passengers were in their 20s. Curiously, fewer were in their early teens.

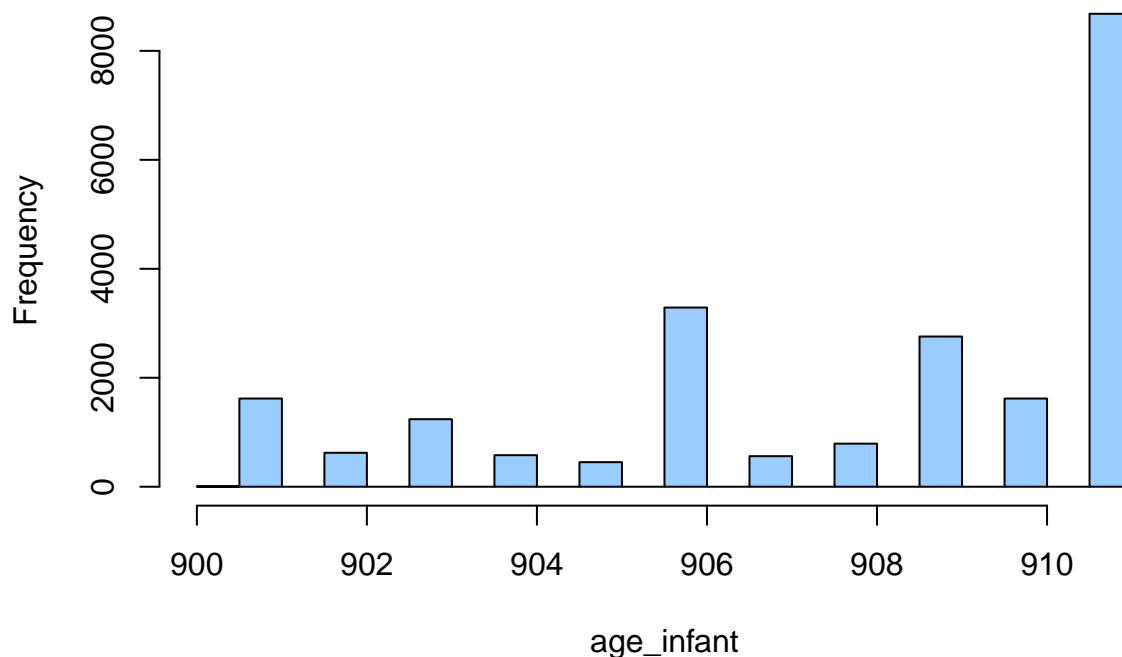
```
age99 <- rus_pass_list$Age[rus_pass_list$Age < 100]
age_infant <- rus_pass_list$Age[rus_pass_list$Age > 800]
hist(age99, main="Histogram of Russian Passenger Ages\n(1 to 99)",
      breaks = seq(1,99,2), col="#99ccff")
```

**Histogram of Russian Passenger Ages  
(1 to 99)**



```
hist(age_infant,  
     main="Histogram of Russian Passenger Ages\n(Infants and Unknown (800))",  
     col="#99ccff")
```

## Histogram of Russian Passenger Ages (Infants and Unknown (800))



## Sex

The sex dataset takes 3 different values (M, F, U).

```
rus_sex      <- read.csv("russian_codes/rus_sex_cl_4530.csv",header=F)
rus_sex
```

```
##   V1      V2 V3
## 1  F  FEMALE NA
## 2  M   MALE  NA
## 3  U UNKNOWN NA
```

```
rus_transit  <- read.csv("russian_codes/rus_transit_cl_4538.csv",header=F)
rus_transit
```

```
##   V1
## 1  C 9
## 2 C09
## 3 C13
## 4  C9
## 5 I09
## 6 I13
## 7 R09
```

```

## 8 R13
## 9 S 9
## 10 S09
## 11 S13
## 12 S15
## 13 S9
## 14 T09
## 15 T13
## 16 T15
##
##
## 1 Citizen of USA [Transit]; Cabin Passenger [Travel]
## 2 Citizen of USA [Transit]; Cabin Passenger [Travel]
## 3 Citizen of USA [Transit]; Steerage [Travel]
## 4 Citizen of USA [Transit]; Cabin Passenger [Travel]
## 5 Final destination not USA or country of origin [Transit]; Cabin Passenger [Travel]
## 6 Final destination not USA or country of origin [Transit]; Steerage [Travel]
## 7 Return trip to USA - non US Citizen [Transit]; Cabin Passenger [Travel]
## 8 Return trip to USA - non US Citizen [Transit]; Steerage [Travel]
## 9 Staying in the USA [Transit]; Cabin Passenger [Travel]
## 10 Staying in the USA [Transit]; Cabin Passenger [Travel]
## 11 Staying in the USA [Transit]; Steerage [Travel]
## 12 Staying in the USA [Transit]; Stowaway [Travel]
## 13 Staying in the USA [Transit]; Cabin Passenger [Travel]
## 14 Returning to country of origin [Transit]; Cabin Passenger [Travel]
## 15 Returning to country of origin [Transit]; Steerage [Travel]
## 16 Returning to country of origin [Transit]; Stowaway [Travel]
##
## V3
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
## 7 NA
## 8 NA
## 9 NA
## 10 NA
## 11 NA
## 12 NA
## 13 NA
## 14 NA
## 15 NA
## 16 NA

```

```

barplot(table(rus_pass_list$Sex), type="b", col="#99ccff")

```

```

## Warning in plot.window(xlim, ylim, log = log, ...): graphical parameter
## "type" is obsolete

```

```

## Warning in axis(if (horiz) 2 else 1, at = at.1, labels = names.arg, lty =
## axis.lty, : graphical parameter "type" is obsolete

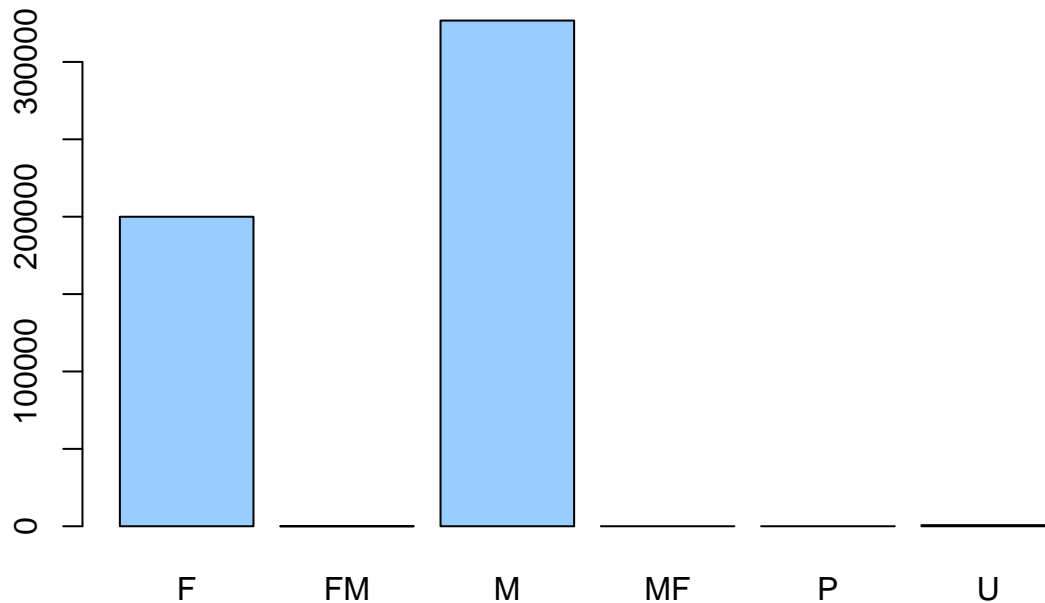
```

```

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## graphical parameter "type" is obsolete

```

```
## Warning in axis(if (horiz) 1 else 2, cex.axis = cex.axis, ...): graphical
## parameter "type" is obsolete
```



## Literacy

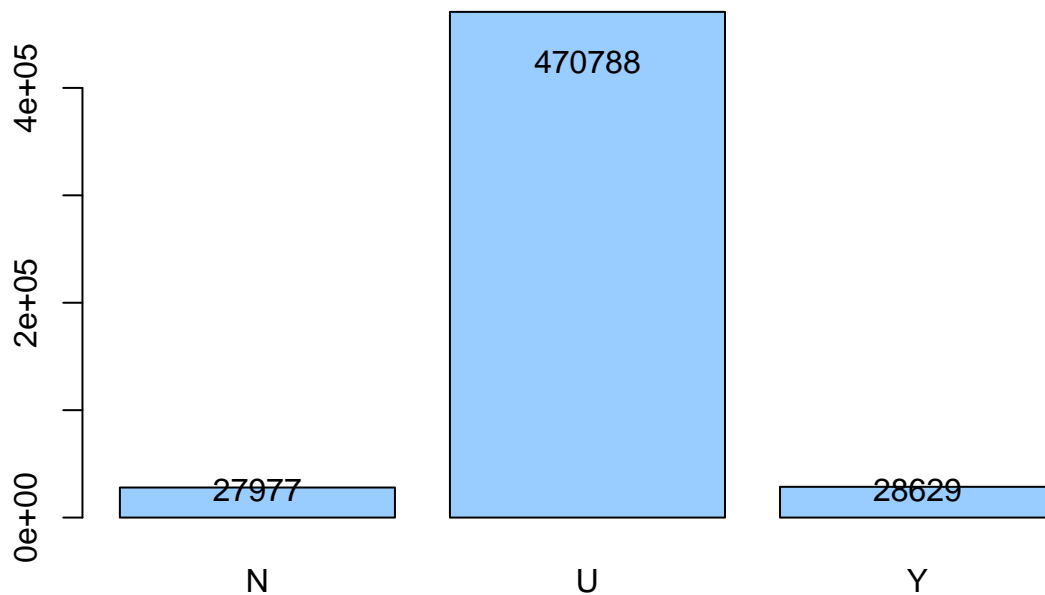
Literacy codes are self explanatory.

```
rus_literacy <- read.csv("russian_codes/rus_literacy_cl_4532.csv",header=F)
rus_literacy
```

```
##   V1          V2 V3
## 1  N          NO NA
## 2  U      UNKNOWN NA
## 3  Y READ & WRITE NA
```

```
t <- table(rus_pass_list$Literacy)
b <- barplot(t, col="#99ccff")
text(b,t*.9,labels=t)
```





## Occupation

Surprisingly, there are 1806 different occupation codes. We may not need to filter out yarn traders, but we can.

```
rus_occupation<- read.csv("russian_codes/rus_occupation_cl_4531.csv",header=F)
head(rus_occupation)
```

```
##      V1      V2 V3
## 1  ABSR  AMBASSADOR NA
## 2  ACCT  ACCOUNTANT NA
## 3  ACHTT  ARCHITECT NA
## 4   ACR    ACROBAT NA
## 5  ACRNP ACORN PICKER NA
## 6   ADJ    ADJUSTER NA
```

```
tail(rus_occupation)
```

```
##      V1      V2 V3
## 1801  YST  YEAST MAKER NA
## 1802  YTRD  YARN TRADER NA
## 1803  YWDR  YARN WINDER NA
## 1804    Z  GRANDCHILD NA
```

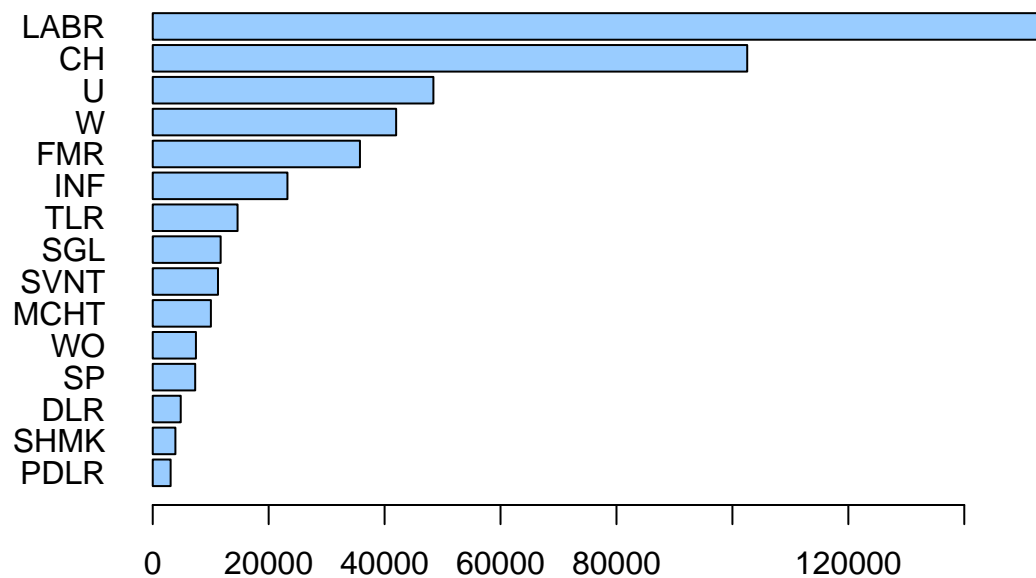
```
## 1805 Z00L ZOOLOGIST NA
## 1806 ZWKR ZINC WORKER NA
```

```
dim(rus_occupation)
```

```
## [1] 1806 3
```

```
occ <- head(sort(table(rus_pass_list$Occupation),decreasing=T),20)
res_known <- rus_pass_list$Occupation[
  rus_pass_list$Occupation!="U"]

par(oma=c(0,2.9,0,0))
barplot(sort(head(occ,15)),
        horiz=T, las=1, col="#99ccff")
```



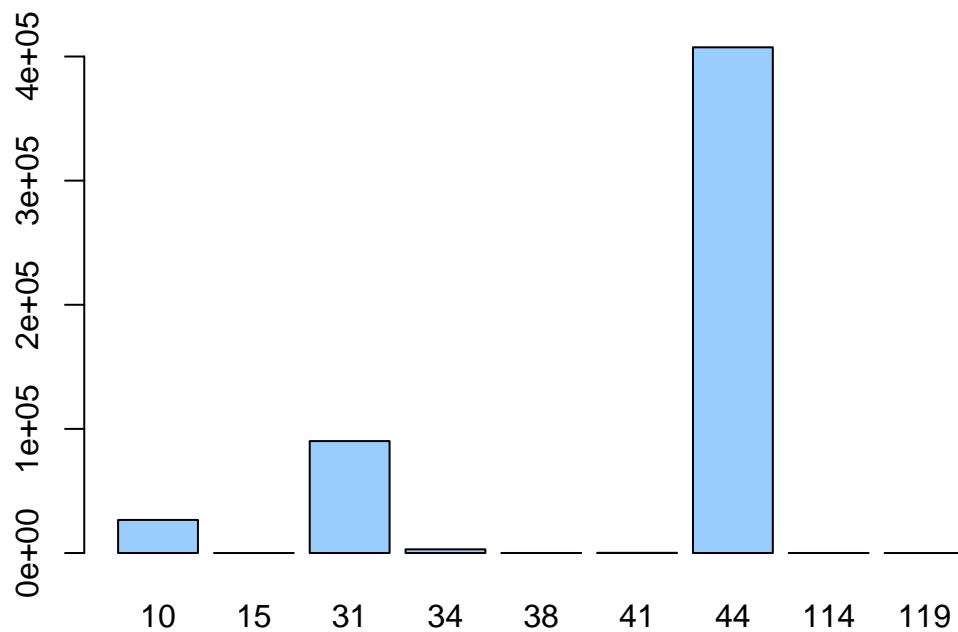
## Country.of.Origin

Below are the origin codes for the Russian passenger lists. Note Galicia, which became part of Poland in more modern times. Also, I believe the “Russian Poles” value refers to Russians of Polish heritage ([https://en.wikipedia.org/wiki/Polish\\_minority\\_in\\_Russia](https://en.wikipedia.org/wiki/Polish_minority_in_Russia)).

```
rus_origin <- read.csv("russian_codes/rus_origin_cl_4671.csv",header=F)
rus_origin
```

```
##      V1      V2 V3
## 1  10    FINLAND NA
## 2  15     RUSSIA NA
## 3  31     POLAND NA
## 4  34    GALICIA NA
## 5  38    ARMENIA NA
## 6  41 RUSSIAN POLES NA
## 7  44     RUSSIA NA
## 8 114     UKRAINE NA
## 9 119    LITHUANIA NA
```

```
barplot(table(rus_pass_list$Country.of.Origin),col="#99ccff")
```



## City.Town.of.Last.Residence

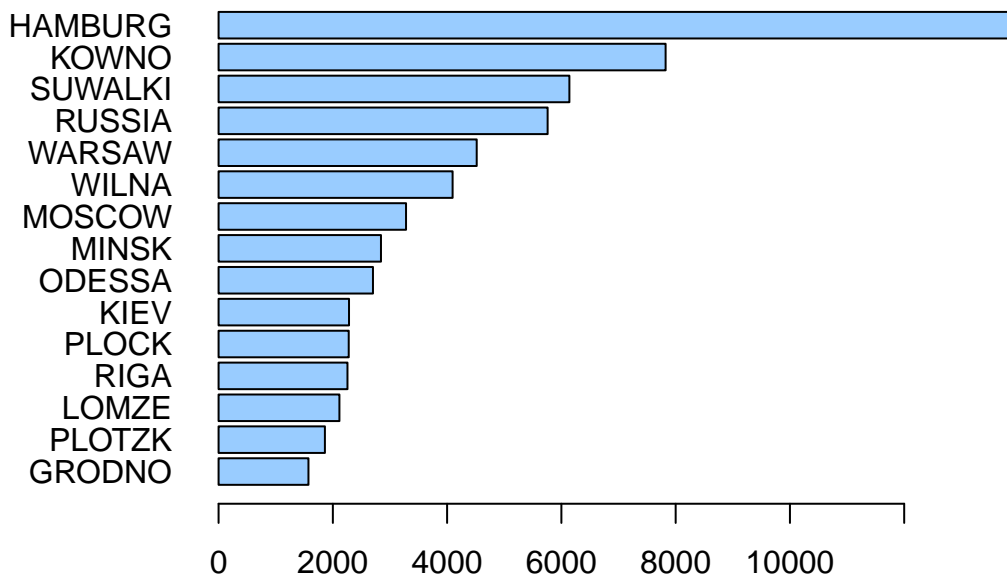
```
head(sort(table(rus_pass_list$City.Town.of.Last.Residence),decreasing=T),20)
```

```
##
##      U      HAMBURG      KOWNO      SUWALKI      RUSSIA
```

```
##      373269      13909      7824      6140      5759
##      WARSAW      WILNA      MOSCOW      MINSK      ODESSA
##      4518      4096      3281      2842      2702
##      KIEV      PLOCK      RIGA      LOMZE      PLOTZK
##      2283      2278      2256      2115      1861
##      GRODNO      BIALYSTOCK ST. PETERBURG      WASA      NEW YORK
##      1573      1553      1498      1175      1117
```

```
res_known <- rus_pass_list$City.Town.of.Last.Residence[
  rus_pass_list$City.Town.of.Last.Residence!="U"]

par(oma=c(0,2.9,0,0))
barplot(tail(sort(table(res_known)),15),
  horiz=T, las=1, col="#99ccff")
```



```
#70 percent of the Last Residence fields are unknown
length(rus_pass_list$City.Town.of.Last.Residence[
  rus_pass_list$City.Town.of.Last.Residence=="U"])/dim(rus_pass_list)[1]
```

```
## [1] 0.7077612
```

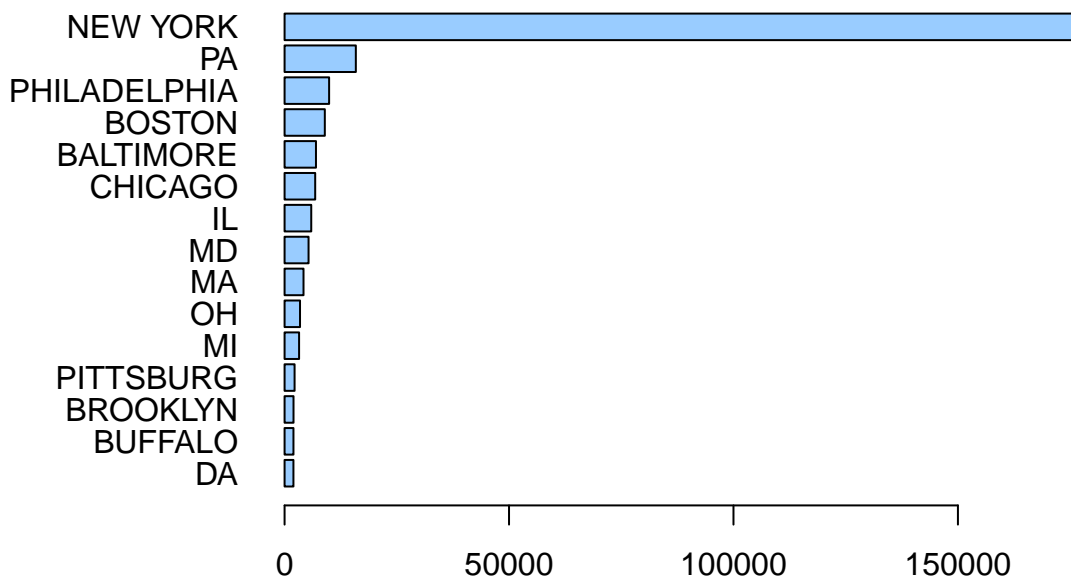
Destination.City.Country

```
head(sort(table(rus_pass_list$Destination.City.Country),decreasing=T),20)
```

```
##
##      USA      NEW YORK      PA PHILADELPHIA      BOSTON
##      228121      177029      15877      9922      8964
##      BALTIMORE      CHICAGO      IL      MD      MA
##      6992      6821      5943      5340      4221
##      OH      MI      PITTSBURG      BROOKLYN      BUFFALO
##      3441      3231      2226      1992      1981
##      DA      NE      WI      MN      KS
##      1972      1629      1534      1378      1238
```

```
usa <- rus_pass_list$Destination.City.Country=="USA"
```

```
par(oma=c(0,2.9,0,0))
barplot(tail(sort(table(rus_pass_list$Destination.City.Country[usa == FALSE])),15),
        horiz=T, las=1, col="#99ccff")
```



```
#Unique US cities
length(unique(rus_pass_list$Destination.City.Country))
```

```
## [1] 2421
```

## Transit.and.or.Travel.Compartment

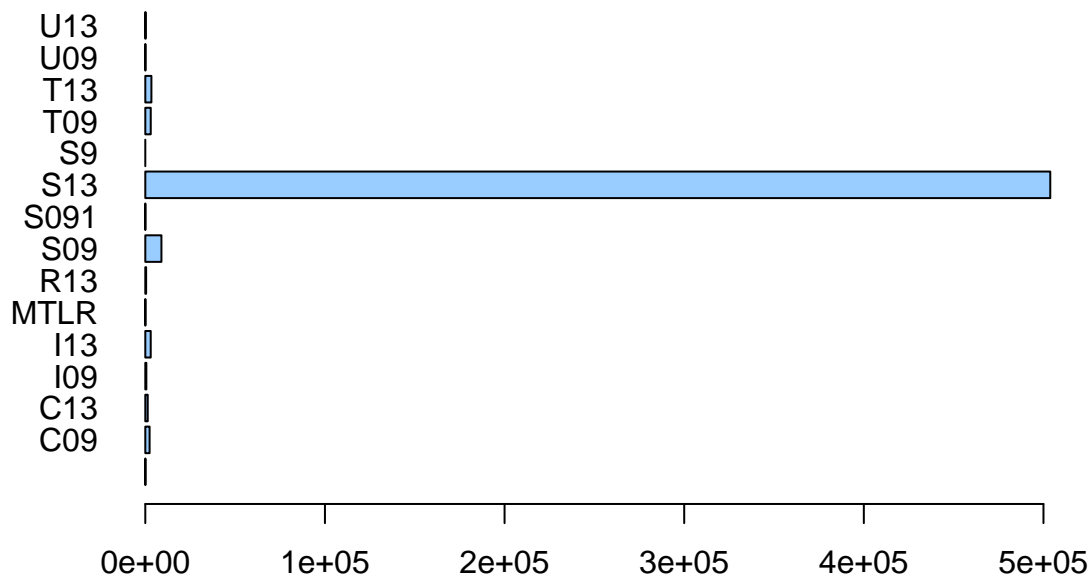
The S13 code is for staying in the US and traveling in steerage.

```
trans <- rus_pass_list$Transit.and.or.Travel.Compartment
```

```
trans_t <- table(trans)
trans_t
```

```
## trans
##      C09  C13  I09  I13  MTLR  R13  S09  S091  S13
##    74 2387 1346  577  3011    64  464  8988    18 503791
##     S9  T09  T13  U09  U13
##     1 3023 3400   68  182
```

```
barplot(trans_t,horiz=T, las=1, col="#99ccff")
```



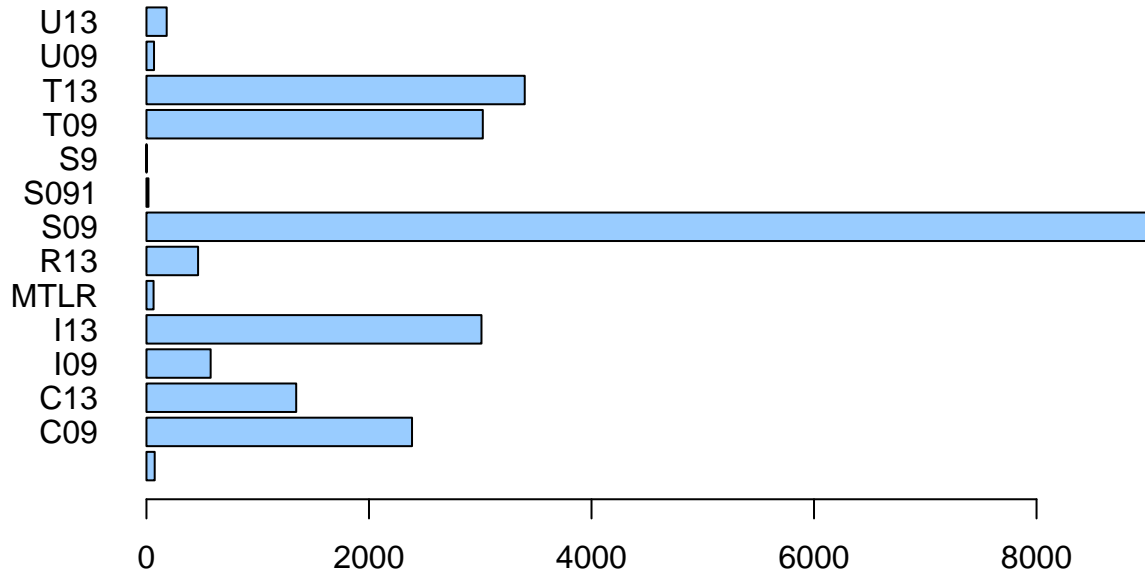
Removing the s13 value allows us to see the distribution for the other codes.

```
#without S13
```

```
trans_t1 <- trans_t[-10]
trans_t1
```

```
## trans
##      C09  C13  I09  I13  MTLR  R13  S09  S091  S9  T09  T13  U09  U13
##    74 2387 1346  577  3011    64  464  8988  18  1 3023 3400   68  182
```

```
barplot(trans_t1,horiz=T, las=1, col="#99ccff")
```



## Big Questions

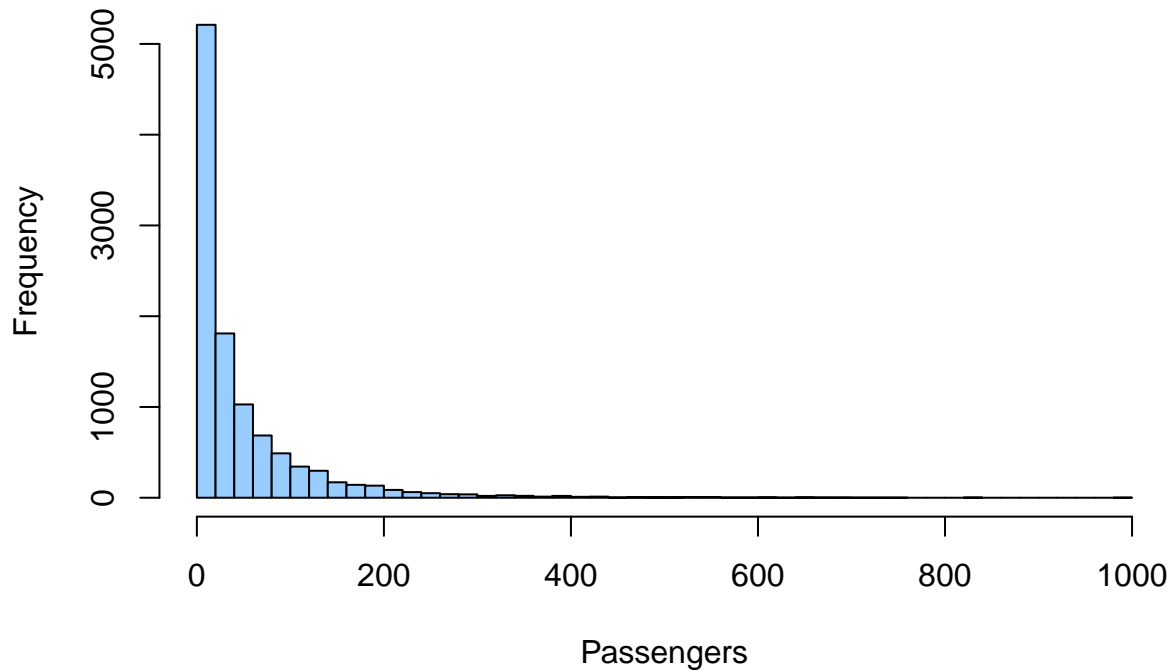
How many passengers were there per ship?

```
ship_nums <- by(rus_pass_list[,1],rus_pass_list[, "Manifest.Identification.Number"],FUN=length)
head(sort(ship_nums,decreasing=))
```

```
## rus_pass_list[, "Manifest.Identification.Number"]
## 459 3447 6431 6492 6535 6539
## 1 1 1 1 1 1
```

```
hist(ship_nums, main="Ship Passenger Numbers", xlab="Passengers", breaks=50,
     col="#99ccff")
```

## Ship Passenger Numbers



What does the time series of the Russian passenger data look like?

To examine this question, I needed to clean the Date.of.Arrival field, so that R will be able to recognize the dates. I did this using the Sublime text editor.

Interestingly, I found that there are some records (~133) with the year as 2012. I made sure that this is there in the original dataset. I'm not sure why they show up here.

```
hdrs_fmt <- read.csv("Data_formatted/manifest_headers_fmt.csv",header=T)
rus_w_head <- merge(hdrs_fmt, rus_pass_list)
by.date <- by(rus_w_head[,1],rus_w_head[, "Date.of.Arrival"],FUN=length)
by.date <- ifelse(is.na(by.date),0,by.date)
```

(As you can see, I'm still working on this question...)