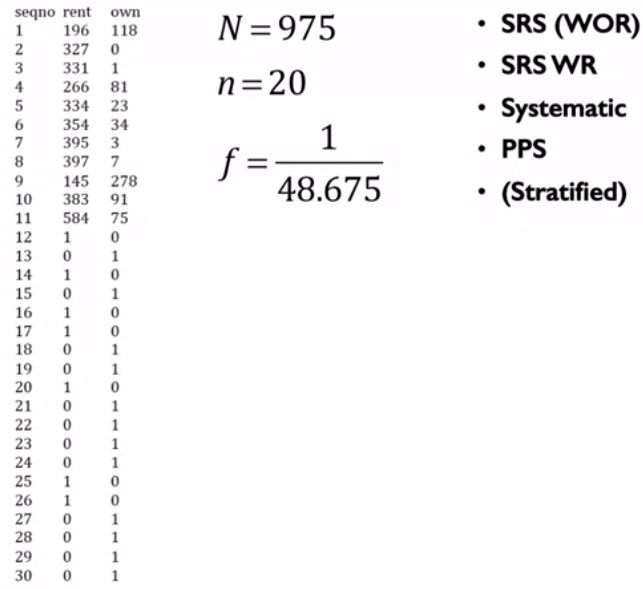
**W6 Pulling it All Together**

**6.1 What about using statistical software to select samples?**

Frame

e.g. Census Blocks



Using R

• Change Directory

setwd(“M:\\Coursera sampling methods”)

• Open data file

frame <- read.table(file = “frame.txt”, header=TRUE, sep = “\t”)

• View data in a spreadsheet

edit(frame)

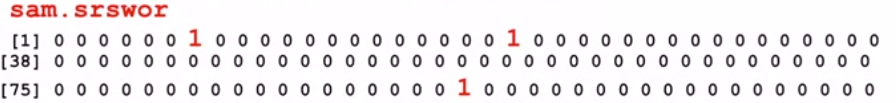
## library(sampling)

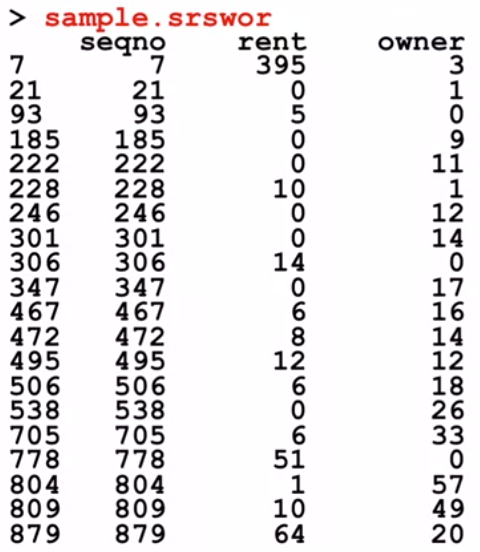
## simple random sample ##

## without replacement ##

sam.srswor <- srswor(n=20, N=975)

sample.srswor <- frame[which(x = (sam.srswor == 1)), ]



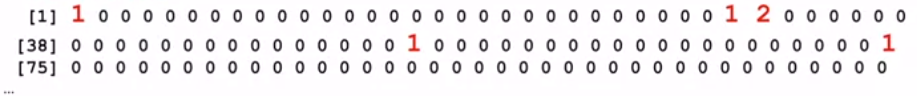


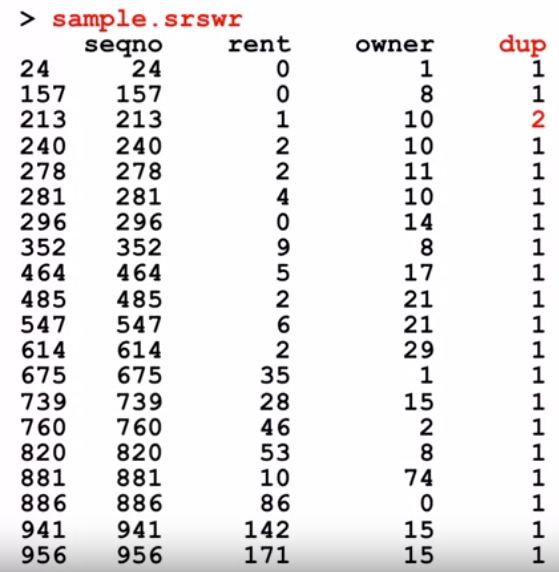
## Simple Random sampling with Replacement ##

sam.srswr <- srswr(n=20, N=975)

sample.srswr <- frame[which(x=(sam.srswr >= 1)), ]

sam.srswr





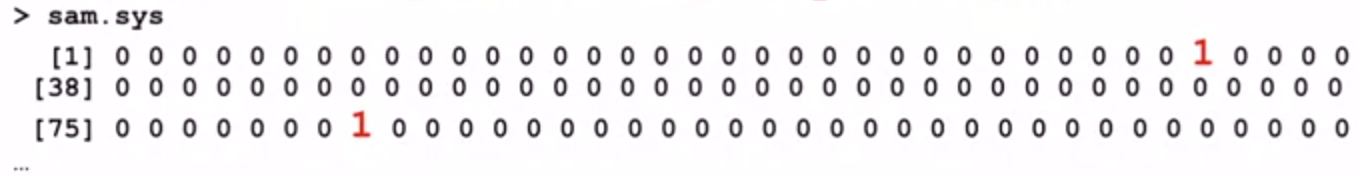
## Systematic Samples ##

## Random start – equal size ##

prob.sys <- rep(x=20/975, times=975)

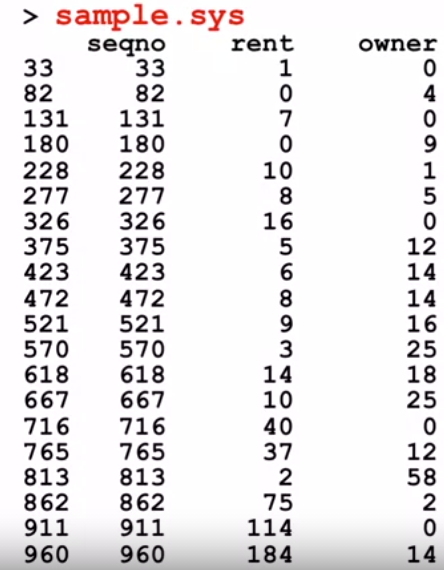
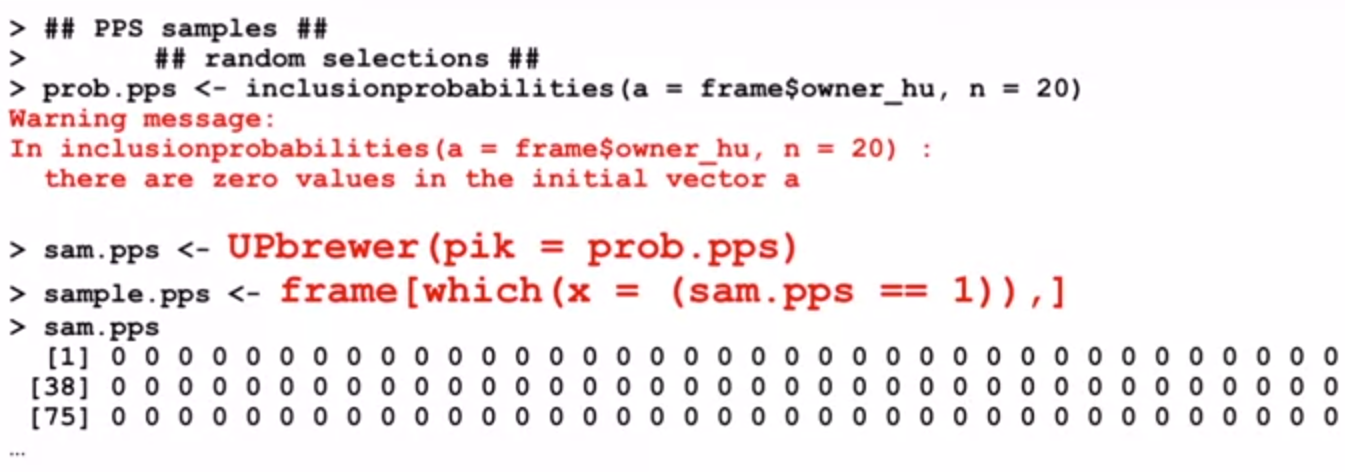
sam.sys <- UPsystematic(pik=prob.sys)

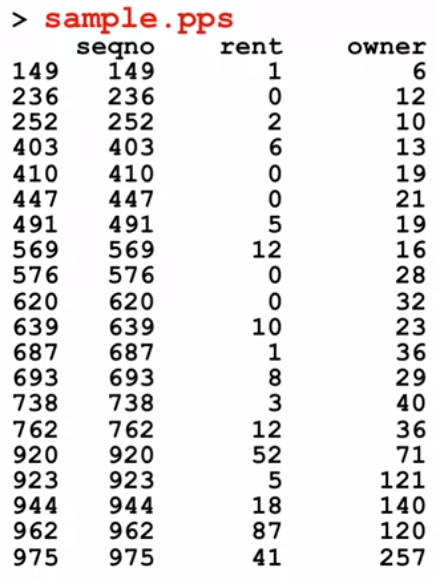
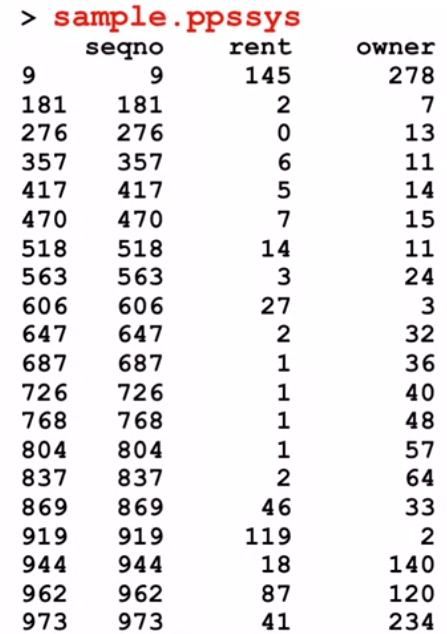
sample.sys <- frame[which(x = (sam.sys == 1)), ]



• Interval = 975/20 = 48.75

• Gaps of 48 and 49

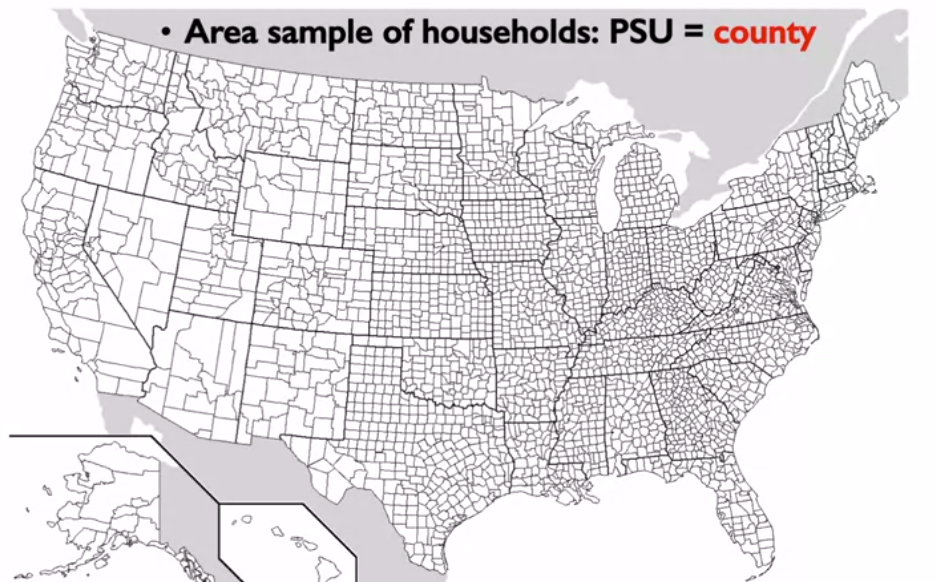
**6.2 Stratified Multistage Sampling**

• First, identify sampling units:

- PSU (Primary Sampling Units)

- Secondary units within PSU, etc.

e.g. Area Sample of households: PSU = county



• Then stratify at each stage

- stratification as a general purpose tool …

- assure representation

- potentially provide gains in precision

• For example, PSU stratification

• Follow principles similar to element stratification

- use cluster characteristics to **create** homogeneous, mutually exclusive, exhaustive **groups**

- stratifying variables, boundaries etc. follow element sampling stratification principles

- **allocate** sample clusters across strata: proportionate, paired, equal, other (Neyman or minimum variance unusual)

- select samples from each group

- later, after data collection, computer estimates separately for each group – statistic and sampling variance

- **combine** results across groups

• Purpose

- control the distribution of the sample

- decrease sampling variance

e.g. Area sample of households: Stratum = MSA(Metropolitan Statistical Areas)

4 categories: very large – medium sized – smaller sized – non metropolitan

e.g. Area sample of households: Stratification = implicit

• Second stage units

Area sample of households: Second stage units

- within selected PSU

- Census tracts across entire selected county?