## Problem Set 3, Question 1

Statistics 506, Fall 2017

```
The RESC data can be downloaded from here.
```

```
# load the libraries
library(tidyverse)
## Warning: package 'tibble' was built under R version 3.4.1
## Warning: package 'purrr' was built under R version 3.4.1
## Warning: package 'dplyr' was built under R version 3.4.1
library(data.table)
# load and clean the dataset
full_data = fread("recs2009_public.csv")
working_data = full_data[,.(REPORTABLE_DOMAIN, ROOFTYPE, NWEIGHT, YEARMADERANGE)
                         ] [ROOFTYPE!=-2]
# convert REPORTABLE_DOMAIN to factor and label the corresponding state name
working_data$state = as.factor(working_data$REPORTABLE_DOMAIN)
levels(working_data$state) = c("CT, ME, NH, RI, VT", "MA", "NY", "NJ", "PA", "IL",
                               "IN, OH", "MI", "WI", "IA, MN, ND, SD", "KS, NE", "MO",
                               "VA", "DE, DC, MD, WV", "GA", "NC, SC", "FL",
                               "AL, KY, MS", "TN", "AR, LA, OK", "TX", "CO",
                               "ID, MT, UT, WY", "AZ", "NV, NM", "CA".
                               "AK, HI, OR, WA" )
# convert ROOFTYPE to factor and label the corresponding type name
working_data$roof_type = as.factor(working_data$ROOFTYPE)
levels(working_data$roof_type) = c("Ceramic or Clay", "Wood Shingles", "Metal",
                                   "Slate", "Composition", "Asphalt", "Concrete",
                                   "Other")
# convert YEARMADERANGE to factor and label accordingly
working_data$decade = as.factor(working_data$YEARMADERANGE)
levels(working_data$decade) = c("pre1950", "1950s", "1960s", "1970s", "1980s", "1990s",
                                "2000s", "2000s")
```

Which state has the highest proportion of wood shingle roofs? Which state(s) the lowest?

The proportion of roof type r in for state s is calculated as:

 $\frac{\text{weighted sum of roof type } r \text{ in state } s}{\text{weighted sum in state } s}$ 

## knitr::kable(wood\_proportion, digits=1,

```
caption="Propotion of roof types by state",
col.names=c("State(s)", "% Wood Shingle Roofs"))
```

State(s)	% Wood Shingle Roofs
NC, SC	14.7
CA	11.0
NV, NM	10.9
CO	10.6
ID, MT, UT, WY	9.8
TX	9.3
FL	8.1
AK, HI, OR, WA	7.5
IN, OH	6.8
DE, DC, MD, WV	6.8
PA	6.7
GA	6.6
AZ	6.1
MA	6.1
NY	5.9
MO	5.7
KS, NE	5.2
CT, ME, NH, RI, VT	4.8
IA, MN, ND, SD	4.6
IL	4.3
VA	3.8
WI	3.8
AR, LA, OK	3.5
AL, KY, MS	3.4
MI	3.4
NJ	2.9
TN	1.8

Table 1: Proposition of roof types by state

From the results, we see that North/South Carolina has the highest percentage of wood shingle roofs, while Tennessee has the lowest.

## Compute the proportion of each roof type for all houses constructed in each decade. Which roof type saw the largest relative rise in use between 1950 and 2000?

Similarly, the proportion of roof type r in decade d is calculated as:

```
}, by=decade] %>%
```

```
## The roof type with the largest relative rise is Ceramic or Clay,
## with relative change 5.5.
```

Roof Type	pre1950	1950s	1960s	1970s	1980s	1990s	2000s	Relative Change
Ceramic or Clay	1.2	1.1	2.0	3.2	5.0	6.6	5.7	5.5
Concrete	0.4	0.8	0.7	1.3	1.4	1.9	3.0	3.5
Composition	54.0	61.2	58.9	56.0	52.1	55.9	64.1	1.0
Metal	6.9	4.5	6.3	13.1	13.8	11.6	4.5	1.0
Wood Shingles	5.8	8.0	8.2	6.3	8.4	6.1	6.4	0.8
Asphalt	26.7	21.4	21.2	17.7	17.4	16.3	15.5	0.7
Slate	2.7	1.6	1.3	0.8	1.2	1.0	0.4	0.3
Other	2.4	1.4	1.3	1.7	0.7	0.6	0.4	0.3

Table 2: Propotion of roof types by decades