CHIC601 Group Project

greyhypotheses

30/11/2021

Exploratory Data Analysis

After inspecting/preparing Social Contact Survey data, in-line with Dr. Read's data preparation suggestions, the outline is

```
'data.frame': 4217 obs. of  14 variables:
 $ id                 : int  1 2 3 4 5 6 7 8 ...
 $ postal             : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 ...
 $ unmatched_postcode : Factor w/ 3 levels "no","yes","unspecified": 1 1 1 1 1 1 1 1 ...
 $ web                : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 ...
 $ age                : int  51 62 36 27 35 61 41 73 ...
 $ date               : Date, format: "2009-05-28" "2009-05-28" ...
 $ day_of_week        : Ord.factor w/ 7 levels "Monday"<"Tuesday"<..: 5 5 5 5 5 5 5 5 ...
 $ postcode           : chr  "KT11 2JF" NA "GU34 2BG" ...
 $ sex                : Factor w/ 3 levels "female","male",..: 1 2 2 1 1 1 1 1 ...
 $ household_size     : Factor w/ 7 levels "2","1","3","4",..: 3 5 3 2 4 2 3 2 ...
 $ occupation         : Factor w/ 17 levels "retired","office",..: 3 10 4 2 7 1 9 1 ...
 $ total_contacts     : int  106 20 7 13 44 30 16 1 ...
 $ method             : Factor w/ 2 levels "postal","online": 1 1 1 1 1 1 1 1 1 ...
 $ agegroup           : Factor w/ 20 levels "[0,5)","[5,10)",..: 11 13 8 6 8 13 9 15 ...
```

The Variables

Age Groups & Survey Method

Disaggregation of responses by survey method

Table 1: The number of survey reponses per survey method

method	frequency
postal	3091
online	1126
NA	0

The age groups, based on the *age* field. [ref. SurveyData() function]

Figure 1: The distribution of age groups
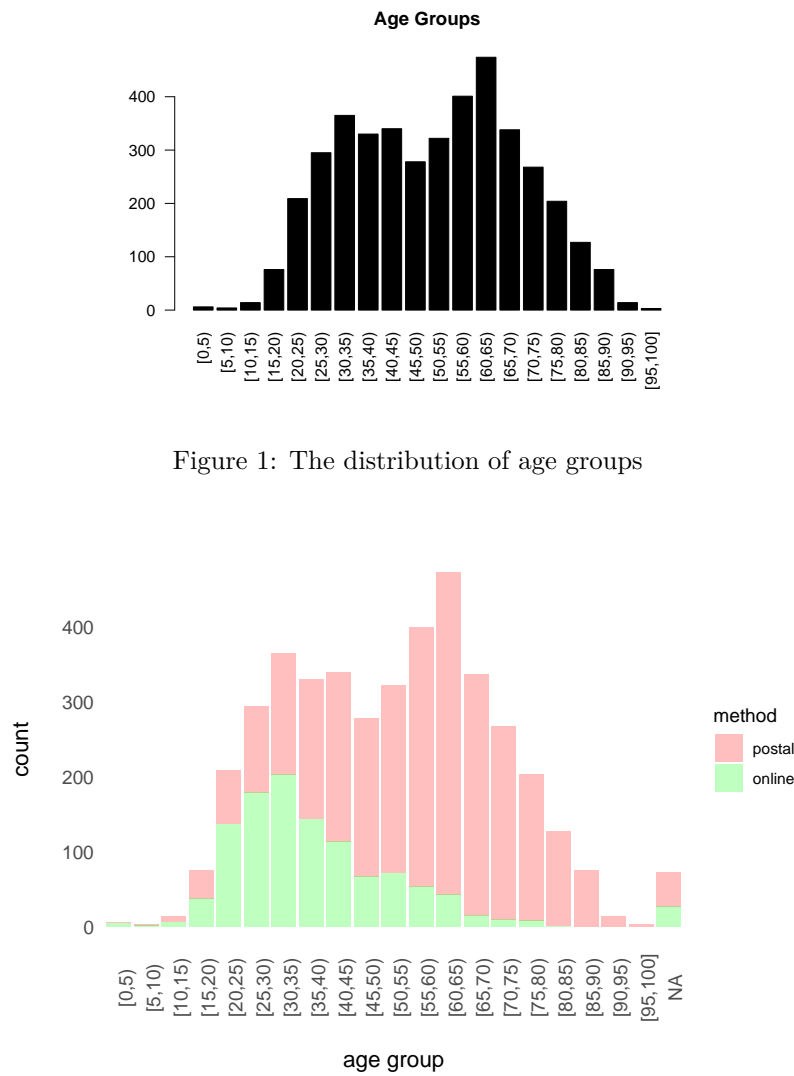
Figure 2: Age group and survey method
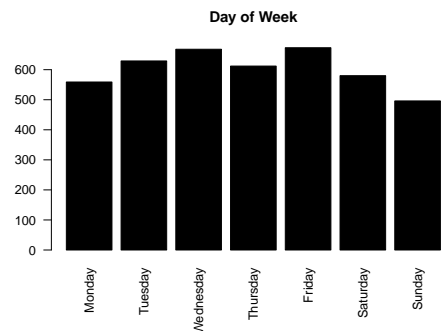
Day of Week & Survey Method

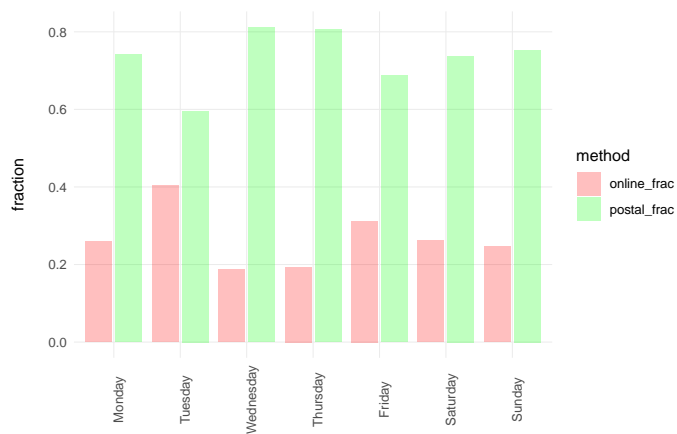Figure 3: The distribution of responses by day of week

Figure 4: The distribution of responses by day of week and survey method
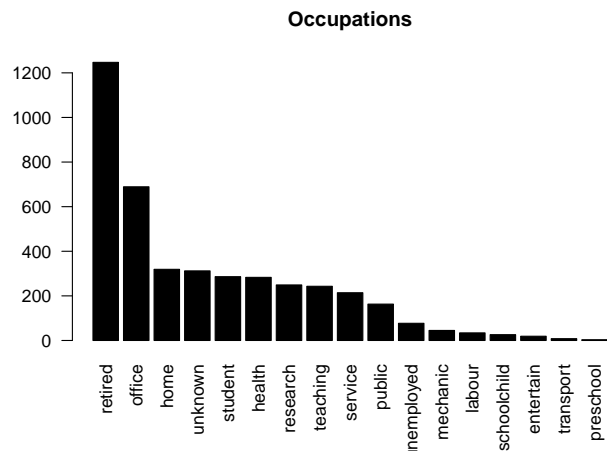
Occupations

Figure 5: The distribution of occupations

Age Group & Sex

Table 2: The sex distribution

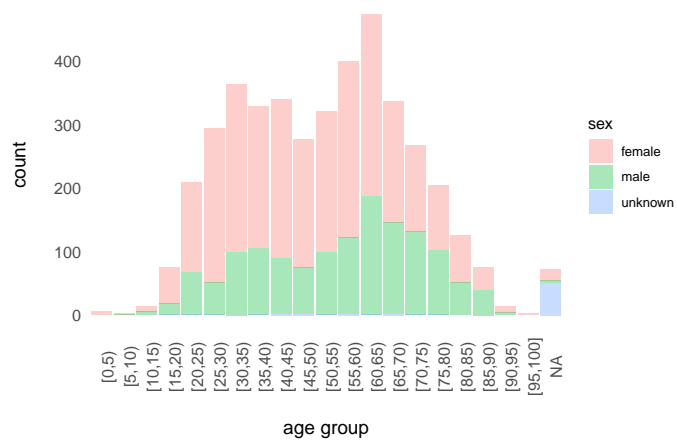| sex	frequency
female	2757
male	1393
unknown	67
NA	0

Figure 6: Age group and sex