WSP setup and demographics code

Lizzie Jones[[1]](#footnote-20)

02/05/2021

# WSP - Initial data exploration

#### About R Markdowns

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>. To generate the document of all content, click the **Knit** button. To change the output (e.g. PDF, HTML) change the ‘output’ at the top to any of the outputs listerd here: <https://rmarkdown.rstudio.com/lesson-9.html>.

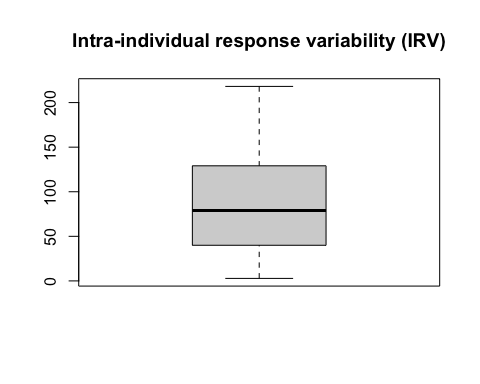
## Data cleaning

First I am going to go through some data cleaning and outline any participants or data points that are removed and explain why. I am using the ‘careless’ pakage first to find evidence of straightlining and

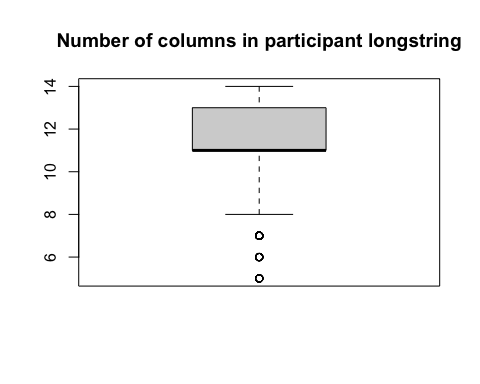
# Average time taken to   
# Percentiles of time   
  
  
  
  
# Data cleaning to investigate straightlining and non-serious participants  
head(colnames(all\_data), 15)

## [1] "SurveyType" "UniqueID\_long" "UniqueID\_short"   
## [4] "TimeTaken" "StartDate" "StartTime"   
## [7] "CompletionDate" "CompletionTime" "Q1\_aware\_stork"   
## [10] "Q2\_photo\_recog" "Q2\_photo\_recog\_score" "Q3\_is\_native"   
## [13] "Q3\_is\_native\_explain" "Q4.1\_migrate" "Q4.1\_migrate\_score"

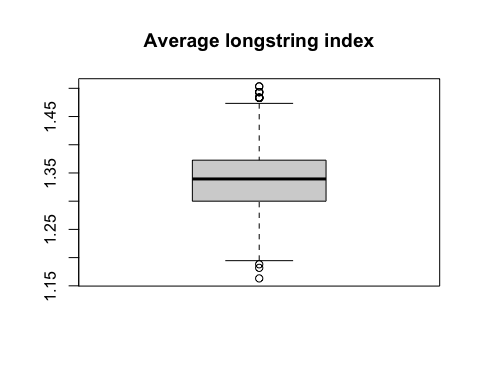
# Calculates the even-odd consistency score  
careless\_all <- evenodd(all\_data, rep(5,10))  
careless\_alldiag <- evenodd(all\_data, rep(5,10), diag = TRUE)  
  
# Calculates the intra-individual response variability (IRV)  
# calculate the irv over all items  
irv\_total <- irv(all\_data)  
boxplot(irv\_total, main="Intra-individual response variability (IRV)")



#calculate the irv over all items + calculate the irv for each quarter of the questionnaire  
irv\_split <- irv(all\_data, split = TRUE, num.split = 4)  
# boxplot(irv\_split$irv4) #produce a boxplot of the IRV for the fourth quarter  
  
# Identifies the longest string of identical consecutive responses for each observation  
careless\_long <- longstring(all\_data, avg = FALSE)  
careless\_avg <- longstring(all\_data, avg = TRUE)  
boxplot(careless\_avg$longstr, main="Number of columns in participant longstring") #produce a boxplot of the longstring index



boxplot(careless\_avg$avgstr, main="Average longstring index")



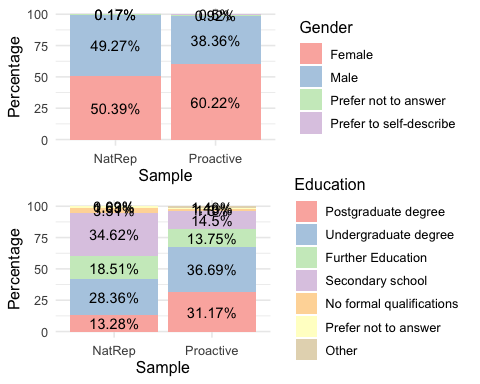
# Straightlining in relation to time taken  
  
  
  
# Find and graph Mahalanobis Distance (D) and flag potential outliers  
## Need to select only numeric/integer columns  
# mahad\_raw <- mahad(all\_data) #only the distances themselves  
# mahad\_flags <- mahad(all\_data, flag = TRUE) #additionally flag outliers  
# mahad\_flags <- mahad(all\_data, flag = TRUE, confidence = 0.999) #Apply a strict criterion

## Exploring participant demographics

The distribution of gender and education is explored and compared between samples using stacked bar plots.

## pct.males pct.female pct.no\_answer pct.self\_desc  
## 1 41.9382 56.99438 0.6741573 0.3932584

## pct.furthered pct.postgrad pct.undergrad pct.secondary pct.no\_formal  
## 1 15.30899 25.30899 33.96067 21.09551 2.022472  
## pct.other pct.not\_answer  
## 1 1.011236 1.292135



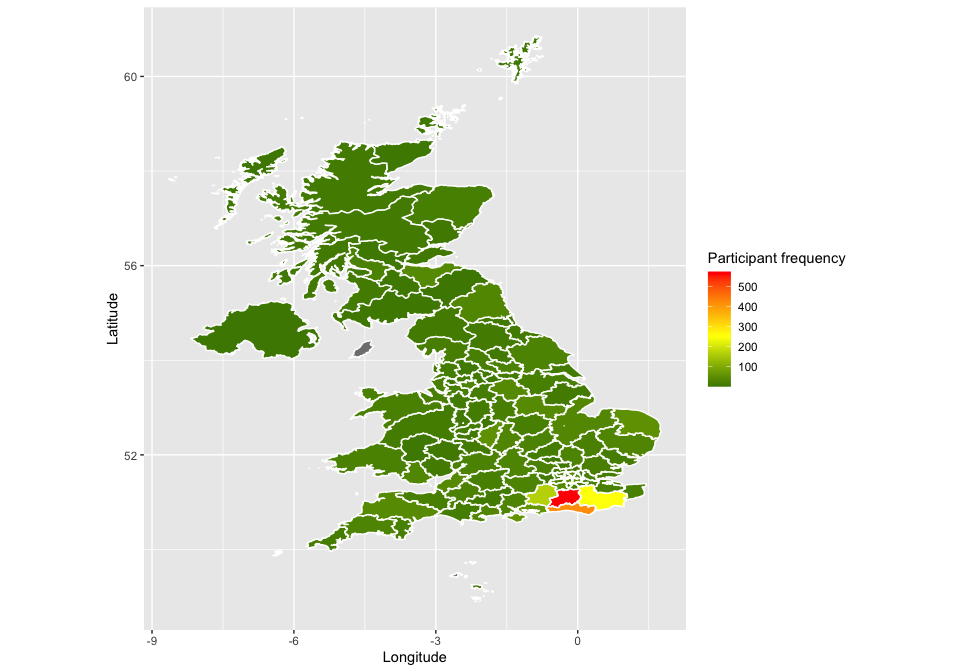
### Participant demographics table

The table below (created using the package “table1”) outlines the demographic characteriscs of each of the two samples, and the overall demographics of all participants across both samples. For each demographic variable the tables provides a breakdown of the number of participants within each level/group and the percentage.

## [1] "<table class=\"Rtable1\">\n<thead>\n<tr>\n<th class='rowlabel firstrow lastrow'></th>\n<th class='firstrow lastrow'><span class='stratlabel'>Nationally rep.<br><span class='stratn'>(N=1167)</span></span></th>\n<th class='firstrow lastrow'><span class='stratlabel'>Proactive<br><span class='stratn'>(N=2393)</span></span></th>\n<th class='firstrow lastrow'><span class='stratlabel'>Overall<br><span class='stratn'>(N=3560)</span></span></th>\n</tr>\n</thead>\n<tbody>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Age group</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>18-24</td>\n<td>126 (10.8%)</td>\n<td>134 (5.6%)</td>\n<td>260 (7.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>25-34</td>\n<td>177 (15.2%)</td>\n<td>333 (13.9%)</td>\n<td>510 (14.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>35-44</td>\n<td>188 (16.1%)</td>\n<td>397 (16.6%)</td>\n<td>585 (16.4%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>45-54</td>\n<td>211 (18.1%)</td>\n<td>489 (20.4%)</td>\n<td>700 (19.7%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>55-64</td>\n<td>195 (16.7%)</td>\n<td>579 (24.2%)</td>\n<td>774 (21.7%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>65-74</td>\n<td>0 (0%)</td>\n<td>368 (15.4%)</td>\n<td>368 (10.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>65+</td>\n<td>270 (23.1%)</td>\n<td>0 (0%)</td>\n<td>270 (7.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>75 and over</td>\n<td>0 (0%)</td>\n<td>81 (3.4%)</td>\n<td>81 (2.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Prefer not to answer</td>\n<td class='lastrow'>0 (0%)</td>\n<td class='lastrow'>12 (0.5%)</td>\n<td class='lastrow'>12 (0.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Gender</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>Female</td>\n<td>588 (50.4%)</td>\n<td>1441 (60.2%)</td>\n<td>2029 (57.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Male</td>\n<td>575 (49.3%)</td>\n<td>918 (38.4%)</td>\n<td>1493 (41.9%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Prefer not to answer</td>\n<td>2 (0.2%)</td>\n<td>22 (0.9%)</td>\n<td>24 (0.7%)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Prefer to self-describe</td>\n<td class='lastrow'>2 (0.2%)</td>\n<td class='lastrow'>12 (0.5%)</td>\n<td class='lastrow'>14 (0.4%)</td>\n</tr>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Education</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>Postgraduate degree</td>\n<td>155 (13.3%)</td>\n<td>746 (31.2%)</td>\n<td>901 (25.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Undergraduate degree</td>\n<td>331 (28.4%)</td>\n<td>878 (36.7%)</td>\n<td>1209 (34.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Further Education</td>\n<td>216 (18.5%)</td>\n<td>329 (13.7%)</td>\n<td>545 (15.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Secondary school</td>\n<td>404 (34.6%)</td>\n<td>347 (14.5%)</td>\n<td>751 (21.1%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>No formal qualifications</td>\n<td>41 (3.5%)</td>\n<td>31 (1.3%)</td>\n<td>72 (2.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Prefer not to answer</td>\n<td>19 (1.6%)</td>\n<td>27 (1.1%)</td>\n<td>46 (1.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Other</td>\n<td class='lastrow'>1 (0.1%)</td>\n<td class='lastrow'>35 (1.5%)</td>\n<td class='lastrow'>36 (1.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Occupation</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>Architecture, Energy & Engineering</td>\n<td>30 (2.6%)</td>\n<td>40 (1.7%)</td>\n<td>70 (2.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Arts, Sport & Media</td>\n<td>21 (1.8%)</td>\n<td>112 (4.7%)</td>\n<td>133 (3.7%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Building & Maintenance</td>\n<td>22 (1.9%)</td>\n<td>22 (0.9%)</td>\n<td>44 (1.2%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Business & Finance</td>\n<td>73 (6.3%)</td>\n<td>128 (5.3%)</td>\n<td>201 (5.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Community & Social Service</td>\n<td>17 (1.5%)</td>\n<td>52 (2.2%)</td>\n<td>69 (1.9%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Computer & Mathematical</td>\n<td>32 (2.7%)</td>\n<td>70 (2.9%)</td>\n<td>102 (2.9%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Education</td>\n<td>67 (5.7%)</td>\n<td>286 (12.0%)</td>\n<td>353 (9.9%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Environment, Nature & Wildlife</td>\n<td>5 (0.4%)</td>\n<td>340 (14.2%)</td>\n<td>345 (9.7%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Farming & Agriculture</td>\n<td>6 (0.5%)</td>\n<td>33 (1.4%)</td>\n<td>39 (1.1%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Fisheries & Aquaculture</td>\n<td>1 (0.1%)</td>\n<td>10 (0.4%)</td>\n<td>11 (0.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Forestry & Woodland Management</td>\n<td>1 (0.1%)</td>\n<td>16 (0.7%)</td>\n<td>17 (0.5%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Healthcare</td>\n<td>88 (7.5%)</td>\n<td>165 (6.9%)</td>\n<td>253 (7.1%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Homemaker / Carer</td>\n<td>16 (1.4%)</td>\n<td>27 (1.1%)</td>\n<td>43 (1.2%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Horticulture/Gardening/Landscaping</td>\n<td>2 (0.2%)</td>\n<td>18 (0.8%)</td>\n<td>20 (0.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Hospitality</td>\n<td>40 (3.4%)</td>\n<td>31 (1.3%)</td>\n<td>71 (2.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Law/Legal</td>\n<td>4 (0.3%)</td>\n<td>12 (0.5%)</td>\n<td>16 (0.4%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Office and Administrative Support</td>\n<td>82 (7.0%)</td>\n<td>120 (5.0%)</td>\n<td>202 (5.7%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Other</td>\n<td>85 (7.3%)</td>\n<td>222 (9.3%)</td>\n<td>307 (8.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Other </td>\n<td>0 (0%)</td>\n<td>1 (0.0%)</td>\n<td>1 (0.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Physical and Social Science</td>\n<td>5 (0.4%)</td>\n<td>18 (0.8%)</td>\n<td>23 (0.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Prefer not to answer</td>\n<td>37 (3.2%)</td>\n<td>47 (2.0%)</td>\n<td>84 (2.4%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Production</td>\n<td>15 (1.3%)</td>\n<td>20 (0.8%)</td>\n<td>35 (1.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Retired</td>\n<td>268 (23.0%)</td>\n<td>382 (16.0%)</td>\n<td>650 (18.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Sales</td>\n<td>51 (4.4%)</td>\n<td>31 (1.3%)</td>\n<td>82 (2.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Student</td>\n<td>54 (4.6%)</td>\n<td>82 (3.4%)</td>\n<td>136 (3.8%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Tourism</td>\n<td>6 (0.5%)</td>\n<td>15 (0.6%)</td>\n<td>21 (0.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Transport</td>\n<td>28 (2.4%)</td>\n<td>30 (1.3%)</td>\n<td>58 (1.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Unemployed</td>\n<td class='lastrow'>111 (9.5%)</td>\n<td class='lastrow'>63 (2.6%)</td>\n<td class='lastrow'>174 (4.9%)</td>\n</tr>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Region</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>East Midlands</td>\n<td>66 (5.7%)</td>\n<td>61 (2.5%)</td>\n<td>127 (3.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>East of England</td>\n<td>100 (8.6%)</td>\n<td>132 (5.5%)</td>\n<td>232 (6.5%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Greater London</td>\n<td>213 (18.3%)</td>\n<td>118 (4.9%)</td>\n<td>331 (9.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>North East</td>\n<td>47 (4.0%)</td>\n<td>29 (1.2%)</td>\n<td>76 (2.1%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>North West</td>\n<td>114 (9.8%)</td>\n<td>61 (2.5%)</td>\n<td>175 (4.9%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Northern Ireland</td>\n<td>0 (0%)</td>\n<td>3 (0.1%)</td>\n<td>3 (0.1%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Scotland</td>\n<td>96 (8.2%)</td>\n<td>56 (2.3%)</td>\n<td>152 (4.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>South East</td>\n<td>174 (14.9%)</td>\n<td>1555 (65.0%)</td>\n<td>1729 (48.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>South West</td>\n<td>104 (8.9%)</td>\n<td>209 (8.7%)</td>\n<td>313 (8.8%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Wales</td>\n<td>58 (5.0%)</td>\n<td>40 (1.7%)</td>\n<td>98 (2.8%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>West Midlands</td>\n<td>106 (9.1%)</td>\n<td>54 (2.3%)</td>\n<td>160 (4.5%)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Yorkshire and the Humber</td>\n<td class='lastrow'>89 (7.6%)</td>\n<td class='lastrow'>75 (3.1%)</td>\n<td class='lastrow'>164 (4.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Area type</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>Rural</td>\n<td>225 (19.3%)</td>\n<td>1047 (43.8%)</td>\n<td>1272 (35.7%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Sub-urban</td>\n<td>548 (47.0%)</td>\n<td>858 (35.9%)</td>\n<td>1406 (39.5%)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Urban</td>\n<td class='lastrow'>394 (33.8%)</td>\n<td class='lastrow'>488 (20.4%)</td>\n<td class='lastrow'>882 (24.8%)</td>\n</tr>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Release site</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>Knepp</td>\n<td>5 (0.4%)</td>\n<td>432 (18.1%)</td>\n<td>437 (12.3%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Knepp-Wintershall</td>\n<td>5 (0.4%)</td>\n<td>265 (11.1%)</td>\n<td>270 (7.6%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>No</td>\n<td>1149 (98.5%)</td>\n<td>1378 (57.6%)</td>\n<td>2527 (71.0%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Wadhurst</td>\n<td>5 (0.4%)</td>\n<td>0 (0%)</td>\n<td>5 (0.1%)</td>\n</tr>\n<tr>\n<td class='rowlabel'>Wadhurst Park</td>\n<td>0 (0%)</td>\n<td>193 (8.1%)</td>\n<td>193 (5.4%)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Wintershall</td>\n<td class='lastrow'>3 (0.3%)</td>\n<td class='lastrow'>125 (5.2%)</td>\n<td class='lastrow'>128 (3.6%)</td>\n</tr>\n</tbody>\n</table>\n"

### Participant postcode mapping

Map so far of participant postcode (first 1 or 2 alphabetical digits, e.g. SW or N)



Participant postcode mapping

1. University of Brighton, [l.jones4@brighton.ac.uk](mailto:l.jones4@brighton.ac.uk) [↑](#footnote-ref-20)