# Survey analysis

## Demographic distribution of participants

What is the gender distribution of responses? We have approximately 4:1 responses from men versus women.

```
chisq.test(table(data$gender))
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(data$gender)
## X-squared = 39.726, df = 1, p-value = 2.923e-10
```

```
summary(data$gender)
```

```
##   Man Woman  NA's
##    90    23     7
```

What is the regional distribution of responses? We balanced the distribution of survey requests across regions, nonetheless developers from some regions where more responsive compared to others. We received at least 10 responses from each region, except Oceania.

```
chisq.test(table(data$regionl))
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(data$regionl)
## X-squared = 46.839, df = 4, p-value = 1.647e-09
```

```
summary(data$regionl)
```

```
##   Europe     Asia Americas   Africa  Oceania     NA's
##       46       29       21       12        4        8
```

For statistical analysis, we selected all regions except Oceania.

```
chisq.test(table(data$regionl1))
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(data$regionl1)
## X-squared = 23.185, df = 3, p-value = 3.695e-05
```

```
summary(data$regionl1)
```

```
##   Europe     Asia Americas   Africa     NA's
##       46       29       21       12       12
```

Region and gender distribution of the respondents

```r
addmargins(table(data[,c("regionl","gender")]))
```

```
##           gender
## regionl    Man Woman Sum
##   Europe    35    10  45
##   Asia      25     4  29
##   Americas  13     7  20
##   Africa    11     1  12
##   Oceania    3     0   3
##   Sum       87    22 109
```

Selection for data analysis

```r
addmargins(table(data[,c("regionl1","gender")]))
```

```
##           gender
## regionl1   Man Woman Sum
##   Europe    35    10  45
##   Asia      25     4  29
##   Americas  13     7  20
##   Africa    11     1  12
##   Sum       84    22 106
```

Is there a pattern in the distribution of men and women across regions? No.

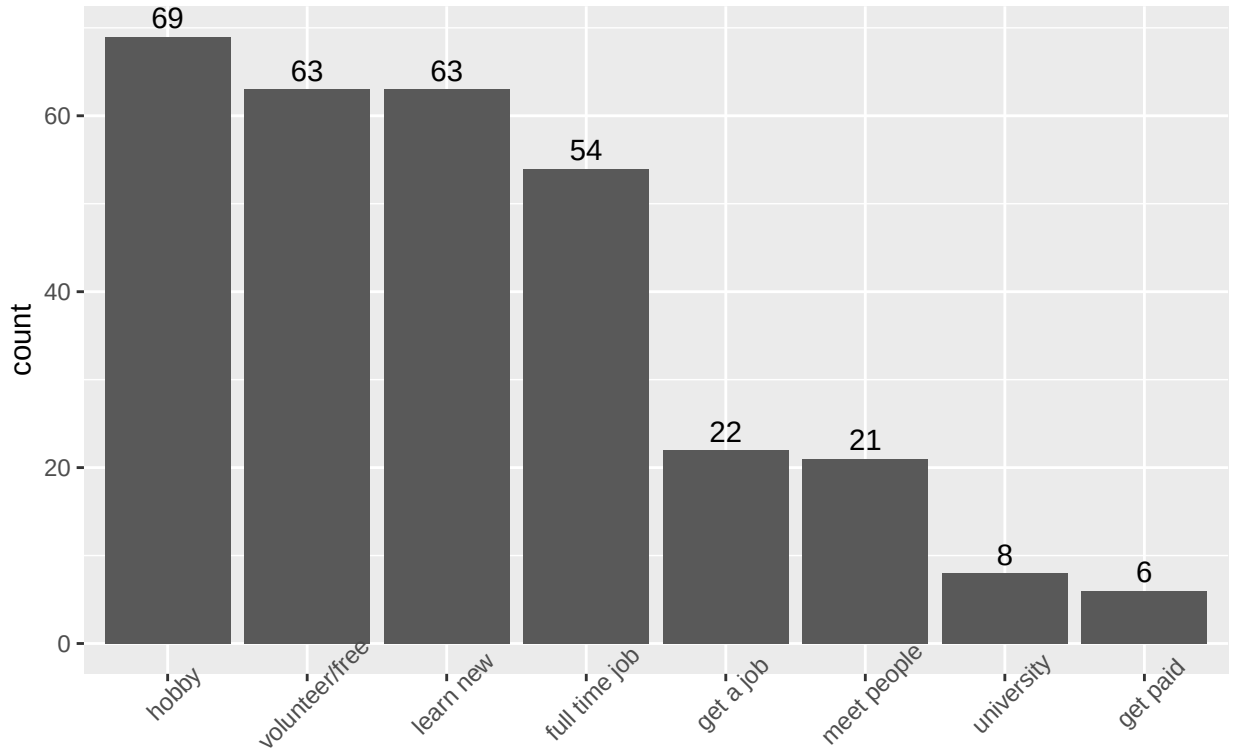
```r
chisq.test(table(data[,c("regionl1","gender")]))
```

```
## Warning in chisq.test(table(data[, c("regionl1", "gender")])): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(data[, c("regionl1", "gender")])
## X-squared = 4.5068, df = 3, p-value = 0.2117
```

## Motivation of developers working in open source software project

Developers primarily work on open source software projects for hobby, volunteer for free, to learn something new, or as a full time job. Other less prominent factors are getting a job, meet people, for university or school and get paid.




Does the motivation vary across gender? No.

```
percentage_distribution(motivation[,c("value","gender")])
```

```
##                                     gender
## value                                Man Woman
##   my full-time job                    16    27
##   my hobby                            24    15
##   volunteer in the community for free  22    13
##   learn something new                 19    27
##   my school or university project      2     5
##   help get a job                       6     7
##   meet new people                      8     5
##   get paid                             3     0
```

```
lapply(1:8,function(i){chisq.test(percentage_distribution(motivation[,c("value","gender")])[i,])})
```

```
## Warning in chisq.test(percentage_distribution(motivation[, c("value",
## "gender")])[i, : Chi-squared approximation may be incorrect

## Warning in chisq.test(percentage_distribution(motivation[, c("value",
## "gender")])[i, : Chi-squared approximation may be incorrect

## [[1]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "gender")])[i,     ]
## X-squared = 2.814, df = 1, p-value = 0.09345
##
##
## [[2]]
```

```
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "gender")])[i,     ]
## X-squared = 2.0769, df = 1, p-value = 0.1495
##
##
## [[3]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "gender")])[i,     ]
## X-squared = 2.3143, df = 1, p-value = 0.1282
##
##
## [[4]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "gender")])[i,     ]
## X-squared = 1.3913, df = 1, p-value = 0.2382
##
##
## [[5]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "gender")])[i,     ]
## X-squared = 1.2857, df = 1, p-value = 0.2568
##
##
## [[6]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "gender")])[i,     ]
## X-squared = 0.076923, df = 1, p-value = 0.7815
##
##
## [[7]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "gender")])[i,     ]
## X-squared = 0.69231, df = 1, p-value = 0.4054
##
##
## [[8]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "gender")])[i,     ]
## X-squared = 3, df = 1, p-value = 0.08326
```

```
#lapply(1:8,function(i){chisq.test(table(motivation[,c("value","gender")])[i,])}) # for actual values;
```

Does the motivation vary across regions?

```
lapply(1:8,function(i){chisq.test(percentage_distribution(motivation[,c("value","regionl1")])[i,])})
```

```
## Warning in chisq.test(percentage_distribution(motivation[, c("value",
## "regionl1")])[i, : Chi-squared approximation may be incorrect

## Warning in chisq.test(percentage_distribution(motivation[, c("value",
## "regionl1")])[i, : Chi-squared approximation may be incorrect

## [[1]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "regionl1")])[i,     ]
## X-squared = 12.909, df = 3, p-value = 0.004837
##
##
## [[2]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "regionl1")])[i,     ]
## X-squared = 4.2771, df = 3, p-value = 0.2331
##
##
## [[3]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "regionl1")])[i,     ]
## X-squared = 2.0118, df = 3, p-value = 0.57
##
##
## [[4]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "regionl1")])[i,     ]
## X-squared = 2.8372, df = 3, p-value = 0.4174
##
##
## [[5]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "regionl1")])[i,     ]
## X-squared = 14.091, df = 3, p-value = 0.002784
##
##
## [[6]]
##
##  Chi-squared test for given probabilities
```

```
##
## data:  percentage_distribution(motivation[, c("value", "region1")])[i,     ]
## X-squared = 4.4, df = 3, p-value = 0.2214
##
##
## [[7]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "region1")])[i,     ]
## X-squared = 6.8065, df = 3, p-value = 0.07833
##
##
## [[8]]
##
##  Chi-squared test for given probabilities
##
## data:  percentage_distribution(motivation[, c("value", "region1")])[i,     ]
## X-squared = 3.3333, df = 3, p-value = 0.343
```

```
percentage_distribution(motivation[,c("value","region1")])
```

```
##                                     region1
## value                                Europe Asia Americas Africa
##   my full-time job                       26   11       21      8
##   my hobby                               21   28       15     19
##   volunteer in the community for free    26   20       17     22
##   learn something new                    15   24       25     22
##   my school or university project         2    1        8      0
##   help get a job                          3    8        8     11
##   meet new people                         5    6        6     14
##   get paid                                2    1        0      3
```

## Frequency of contribution

```
data[,"frequency.contribution"]<-ordered(data[,"frequency.contribution"],
                                         c("Hourly","Daily","Weekly","Monthly"))
```

```
table(data[,"frequency.contribution"])
```

```
##
##  Hourly   Daily  Weekly Monthly
##       4      12      22      79
```

```
chisq.test(table(data[,"frequency.contribution"]))
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(data[, "frequency.contribution"])
## X-squared = 118.38, df = 3, p-value < 2.2e-16
```

Does frequency of contribution vary across gender? No.

```
chisq.test(table(data[,c("frequency.contribution","gender")]))
```

```
## Warning in chisq.test(table(data[, c("frequency.contribution", "gender")])):
## Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(data[, c("frequency.contribution", "gender")])
## X-squared = 2.8463, df = 3, p-value = 0.4159
```

```
percentage_distribution(data[,c("frequency.contribution","gender")])
```

```
##                        gender
## frequency.contribution Man Woman
##                Hourly    5     0
##                Daily    10    14
##                Weekly   20     9
##                Monthly  65    77
```

Does frequency of contribution vary across regions? No.

```
chisq.test(table(data[,c("frequency.contribution","regionl1")]))
```

```
## Warning in chisq.test(table(data[, c("frequency.contribution", "regionl1")])):
## Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(data[, c("frequency.contribution", "regionl1")])
## X-squared = 11.621, df = 9, p-value = 0.2355
```

```
percentage_distribution(data[,c("frequency.contribution","regionl1")])
```

```
##                        regionl1
## frequency.contribution Europe Asia Americas Africa
##                Hourly       2    3        0      8
##                Daily        9   17       10      8
##                Weekly      14   31        5     17
##                Monthly     75   48       86     67
```

## Selection of projects

```
chisq.test(table(selection[,"selection.how.software.is.build"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(selection[, "selection.how.software.is.build"])[c(1, 3)]
## X-squared = 31.696, df = 1, p-value = 1.803e-08
```

```
#table(selection[,"selection.how.software.is.build"])
round(prop.table(table(selection[,"selection.how.software.is.build"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
```

```
##           79           21
```

```
chisq.test(table(selection[,"selection.project.goal.align"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(selection[, "selection.project.goal.align"])[c(1, 3)]
## X-squared = 86.627, df = 1, p-value < 2.2e-16
```

```
#table(selection[,"selection.project.goal.align"])
round(prop.table(table(selection[,"selection.project.goal.align"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            96             4
```

```
chisq.test(table(selection[,"selection.friends.contribute"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(selection[, "selection.friends.contribute"])[c(1, 3)]
## X-squared = 9.6667, df = 1, p-value = 0.001876
```

```
#table(selection[,"selection.friends.contribute"])
round(prop.table(table(selection[,"selection.friends.contribute"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            33            67
```

```
chisq.test(table(selection[,"selection.project.welcoming"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(selection[, "selection.project.welcoming"])[c(1, 3)]
## X-squared = 39.13, df = 1, p-value = 3.964e-10
```

```
#table(selection[,"selection.project.welcoming"])
round(prop.table(table(selection[,"selection.project.welcoming"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            83            17
```

```
chisq.test(table(selection[,"selection.easy.to.join"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(selection[, "selection.easy.to.join"])[c(1, 3)]
## X-squared = 34.844, df = 1, p-value = 3.571e-09
```

```
#table(selection[,"selection.easy.to.join"])
round(prop.table(table(selection[,"selection.easy.to.join"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
```

```
##            81            19
```

```
chisq.test(table(selection[,"selection.saw.on.social.media"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(selection[, "selection.saw.on.social.media"])[c(1, 3)]
## X-squared = 66.176, df = 1, p-value = 4.123e-16
```

```
#table(selection[,"selection.saw.on.social.media"])
round(prop.table(table(selection[,"selection.saw.on.social.media"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##             6            94
```

Does the criteria for the selection of projects vary across gender? Partial. Yes: friends.contribute.

```
lapply(1:6,function(X){chisq.test(table(selection[,c(X,7)])[c(1,3),])})
```

```
## [[1]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(selection[, c(X, 7)])[c(1, 3), ]
## X-squared = 2.2411, df = 1, p-value = 0.1344
##
##
## [[2]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(selection[, c(X, 7)])[c(1, 3), ]
## X-squared = 0.17574, df = 1, p-value = 0.6751
##
##
## [[3]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(selection[, c(X, 7)])[c(1, 3), ]
## X-squared = 6.56, df = 1, p-value = 0.01043
##
##
## [[4]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(selection[, c(X, 7)])[c(1, 3), ]
## X-squared = 1.1058, df = 1, p-value = 0.293
##
##
## [[5]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data:  table(selection[, c(X, 7)])[c(1, 3), ]
## X-squared = 0.14429, df = 1, p-value = 0.7041
##
##
## [[6]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(selection[, c(X, 7)])[c(1, 3), ]
## X-squared = 0.10796, df = 1, p-value = 0.7425
```

```
lapply(1:6,function(X){percentage_distribution(selection[,c(X,7)])})
```

```
## [[1]]
##                                gender
## selection.how.software.is.build Man Woman
##                   Important      57    73
##                   Neutral        22    23
##                   Not important  21     5
##
## [[2]]
##                               gender
## selection.project.goal.align Man Woman
##                Important      80    91
##                Neutral        16     9
##                Not important   4     0
##
## [[3]]
##                               gender
## selection.friends.contribute Man Woman
##                Important      20    41
##                Neutral        22    36
##                Not important  59    23
##
## [[4]]
##                              gender
## selection.project.welcoming Man Woman
##               Important      61    68
##               Neutral        22    27
##               Not important  17     5
##
## [[5]]
##                        gender
## selection.easy.to.join Man Woman
##          Important      61    59
##          Neutral        22    32
##          Not important  17     9
##
## [[6]]
##                               gender
## selection.saw.on.social.media Man Woman
##                 Important       5     0
##                 Neutral        25    32
##                 Not important  70    68
```

```r
prop.table(table(selection[,c(3,7)])[c(1,3),],2)*100 # distribution for projects with friends
```

```
##                              gender
## selection.friends.contribute      Man    Woman
##                Important     25.00000 64.28571
##                Not important 75.00000 35.71429
```

Does the criteria for the selection of projects vary across regions? No.

```r
lapply(1:6,function(X){chisq.test(table(selection[,c(X,8)])[c(1,3),])})
```

```
## [[1]]
##
##  Pearson's Chi-squared test
##
## data:  table(selection[, c(X, 8)])[c(1, 3), ]
## X-squared = 1.431, df = 3, p-value = 0.6983
##
##
## [[2]]
##
##  Pearson's Chi-squared test
##
## data:  table(selection[, c(X, 8)])[c(1, 3), ]
## X-squared = 2.8168, df = 3, p-value = 0.4207
##
##
## [[3]]
##
##  Pearson's Chi-squared test
##
## data:  table(selection[, c(X, 8)])[c(1, 3), ]
## X-squared = 0.3214, df = 3, p-value = 0.956
##
##
## [[4]]
##
##  Pearson's Chi-squared test
##
## data:  table(selection[, c(X, 8)])[c(1, 3), ]
## X-squared = 0.93212, df = 3, p-value = 0.8177
##
##
## [[5]]
##
##  Pearson's Chi-squared test
##
## data:  table(selection[, c(X, 8)])[c(1, 3), ]
## X-squared = 4.3409, df = 3, p-value = 0.2269
##
##
## [[6]]
##
##  Pearson's Chi-squared test
##
```

```
## data:  table(selection[, c(X, 8)])[c(1, 3), ]
## X-squared = 1.7178, df = 3, p-value = 0.633
```

```
lapply(1:6,function(X){percentage_distribution(selection[,c(X,8)])})
```

```
## [[1]]
##                                 region1
## selection.how.software.is.build Europe Asia Americas Africa
##                   Important         59   57       60     83
##                   Neutral           20   29       25      8
##                   Not important     20   14       15      8
##
## [[2]]
##                               region1
## selection.project.goal.align Europe Asia Americas Africa
##                Important         89   76       81     75
##                Neutral            9   17       19     17
##                Not important      2    7        0      8
##
## [[3]]
##                               region1
## selection.friends.contribute Europe Asia Americas Africa
##                Important         26   24       24     18
##                Neutral           23   21       33     27
##                Not important     51   55       43     55
##
## [[4]]
##                              region1
## selection.project.welcoming Europe Asia Americas Africa
##               Important         64   62       57     83
##               Neutral           22   21       33      8
##               Not important     13   17       10      8
##
## [[5]]
##                         region1
## selection.easy.to.join Europe Asia Americas Africa
##          Important         56   59       67     82
##          Neutral           22   31       29      9
##          Not important     22   10        5      9
##
## [[6]]
##                                region1
## selection.saw.on.social.media Europe Asia Americas Africa
##                 Important          2    4        0      9
##                 Neutral           20   32       38      9
##                 Not important     77   64       62     82
```

## Continue participation

```
table(continue[,"continue.interaction.with.welcoming.contributors"])
```

```
##
##     Important       Neutral Not important
##            92            16             9
```

```r
chisq.test(table(continue[,"continue.interaction.with.welcoming.contributors"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(continue[, "continue.interaction.with.welcoming.contributors"])[c(1,     3)]
## X-squared = 68.208, df = 1, p-value < 2.2e-16
```

```r
round(prop.table(table(continue[,"continue.interaction.with.welcoming.contributors"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            91              9
```

```r
table(continue[,"continue.connects.with.people.worldwide"])
```

```
##
##     Important       Neutral Not important
##            64            31            18
```

```r
chisq.test(table(continue[,"continue.connects.with.people.worldwide"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(continue[, "continue.connects.with.people.worldwide"])[c(1,     3)]
## X-squared = 25.805, df = 1, p-value = 3.777e-07
```

```r
round(prop.table(table(continue[,"continue.connects.with.people.worldwide"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            78            22
```

```r
#table(continue[,"continue.low.stress.levels"])
chisq.test(table(continue[,"continue.low.stress.levels"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(continue[, "continue.low.stress.levels"])[c(1, 3)]
## X-squared = 20.513, df = 1, p-value = 5.923e-06
```

```r
round(prop.table(table(continue[,"continue.low.stress.levels"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            76            24
```

```r
table(continue[,"continue.exciting.tasks"])
```

```
##
##     Important       Neutral Not important
##            77            26            14
```

```r
chisq.test(table(continue[,"continue.exciting.tasks"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
```

```
## data:  table(continue[, "continue.exciting.tasks"])[c(1, 3)]
## X-squared = 43.615, df = 1, p-value = 3.997e-11
```

```
round(prop.table(table(continue[,"continue.exciting.tasks"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            85            15
```

```
#table(continue[,"continue.challenging.tasks"])
chisq.test(table(continue[,"continue.exciting.tasks"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(continue[, "continue.exciting.tasks"])[c(1, 3)]
## X-squared = 43.615, df = 1, p-value = 3.997e-11
```

```
round(prop.table(table(continue[,"continue.challenging.tasks"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            90            10
```

```
#table(continue[,"continue.being.paid"])
chisq.test(table(continue[,"continue.being.paid"])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(continue[, "continue.being.paid"])[c(1, 3)]
## X-squared = 3.2, df = 1, p-value = 0.07364
```

```
round(prop.table(table(continue[,"continue.being.paid"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            40            60
```

Does the criteria for continued participation in projects vary across gender? Partial. Continue.being.paid

```
lapply(1:6,function(X){chisq.test(percentage_distribution(continue[,c(X,7)])[c(1,3),])})
```

```
## [[1]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(continue[, c(X, 7)])[c(1, 3), ]
## X-squared = 1.0902, df = 1, p-value = 0.2964
##
##
## [[2]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(continue[, c(X, 7)])[c(1, 3), ]
## X-squared = 4.3247e-31, df = 1, p-value = 1
##
##
```

```
## [[3]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(continue[, c(X, 7)])[c(1, 3), ]
## X-squared = 2.4373, df = 1, p-value = 0.1185
##
##
## [[4]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(continue[, c(X, 7)])[c(1, 3), ]
## X-squared = 0.49317, df = 1, p-value = 0.4825
##
##
## [[5]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(continue[, c(X, 7)])[c(1, 3), ]
## X-squared = 1.287, df = 1, p-value = 0.2566
##
##
## [[6]]
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(continue[, c(X, 7)])[c(1, 3), ]
## X-squared = 9.9847, df = 1, p-value = 0.001578
```

```
lapply(1:6,function(X){percentage_distribution(continue[,c(X,7)])})
```

```
## [[1]]
##                                                gender
## continue.interaction.with.welcoming.contributors Man Woman
##                                     Important      78    78
##                                     Neutral        12    17
##                                     Not important   9     4
##
## [[2]]
##                                          gender
## continue.connects.with.people.worldwide Man Woman
##                            Important      56    61
##                            Neutral        29    22
##                            Not important  15    17
##
## [[3]]
##                            gender
## continue.low.stress.levels Man Woman
##              Important      53    39
##              Neutral        31    39
##              Not important  15    22
##
## [[4]]
```

```
##                         gender
## continue.exciting.tasks Man Woman
##           Important      66    65
##           Neutral        20    26
##           Not important  14     9
##
## [[5]]
##                            gender
## continue.challenging.tasks Man Woman
##              Important      70    74
##              Neutral        21    22
##              Not important   9     4
##
## [[6]]
##                     gender
## continue.being.paid Man Woman
##       Important      26    39
##       Neutral        26    39
##       Not important  48    22
```

```
round(prop.table(table(continue[,c(6,7)])[c(1,3),],2)*100,0) # criteria to continue participation: bein
```

```
##                     gender
## continue.being.paid Man Woman
##       Important      35    64
##       Not important  65    36
```

Does the criteria for continued participation in projects vary across regions? Yes.

```
lapply(1:6,function(X){chisq.test(percentage_distribution(continue[,c(X,8)])[c(1,3),])})
```

```
## Warning in chisq.test(percentage_distribution(continue[, c(X, 8)])[c(1, : Chi-
## squared approximation may be incorrect
```

```
## [[1]]
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(continue[, c(X, 8)])[c(1, 3), ]
## X-squared = 13.633, df = 3, p-value = 0.003449
##
##
## [[2]]
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(continue[, c(X, 8)])[c(1, 3), ]
## X-squared = 31.954, df = 3, p-value = 5.351e-07
##
##
## [[3]]
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(continue[, c(X, 8)])[c(1, 3), ]
## X-squared = 12.285, df = 3, p-value = 0.006468
```

```
##
##
## [[4]]
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(continue[, c(X, 8)])[c(1, 3), ]
## X-squared = 29.842, df = 3, p-value = 1.49e-06
##
##
## [[5]]
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(continue[, c(X, 8)])[c(1, 3), ]
## X-squared = 33.037, df = 3, p-value = 3.164e-07
##
##
## [[6]]
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(continue[, c(X, 8)])[c(1, 3), ]
## X-squared = 35.206, df = 3, p-value = 1.102e-07
```

```
lapply(1:6,function(X){percentage_distribution(continue[,c(X,8)])})
```

```
## [[1]]
##                                                 region1
## continue.interaction.with.welcoming.contributors Europe Asia Americas Africa
##                                     Important         71   93       81     75
##                                     Neutral           18    4       14     25
##                                     Not important     11    4        5      0
##
## [[2]]
##                                       region1
## continue.connects.with.people.worldwide Europe Asia Americas Africa
##                            Important         45   62       67     73
##                            Neutral           32   31       14     27
##                            Not important     23    8       19      0
##
## [[3]]
##                            region1
## continue.low.stress.levels Europe Asia Americas Africa
##              Important         47   63       48     55
##              Neutral           31   30       38     36
##              Not important     22    7       14      9
##
## [[4]]
##                         region1
## continue.exciting.tasks Europe Asia Americas Africa
##           Important         55   86       48     92
##           Neutral           27   14       38      0
##           Not important     18    0       14      8
##
```

```
## [[5]]
##                            regionl1
## continue.challenging.tasks Europe Asia Americas Africa
##              Important         61   86       67    100
##              Neutral           27   14       19      0
##              Not important     11    0       14      0
##
## [[6]]
##                    regionl1
## continue.being.paid Europe Asia Americas Africa
##       Important         28   22       15     42
##       Neutral           19   41       30     42
##       Not important     53   37       55     17
```

```
lapply(1:6,function(i){round(prop.table(table(continue[,c(i,8)])[c(1,3),],2)*100,0)})
```

```
## [[1]]
##                                                 regionl1
## continue.interaction.with.welcoming.contributors Europe Asia Americas Africa
##                                   Important         86   96       94    100
##                                   Not important     14    4        6      0
##
## [[2]]
##                                       regionl1
## continue.connects.with.people.worldwide Europe Asia Americas Africa
##                           Important         67   89       78    100
##                           Not important     33   11       22      0
##
## [[3]]
##                            regionl1
## continue.low.stress.levels Europe Asia Americas Africa
##              Important         68   89       77     86
##              Not important     32   11       23     14
##
## [[4]]
##                         regionl1
## continue.exciting.tasks Europe Asia Americas Africa
##           Important         75  100       77     92
##           Not important     25    0       23      8
##
## [[5]]
##                            regionl1
## continue.challenging.tasks Europe Asia Americas Africa
##              Important         84  100       82    100
##              Not important     16    0       18      0
##
## [[6]]
##                    regionl1
## continue.being.paid Europe Asia Americas Africa
##       Important         34   38       21     71
##       Not important     66   62       79     29
```

### Importance of same geographic region

Same geographic region is not important.

```
round(prop.table(table(data[,"important.same.geographic.region"]))*100,0)
```

```
##
##     Important       Neutral Not important
##            13            26            61
```

```
chisq.test(round(prop.table(table(data[,"important.same.geographic.region"]))*100,0)[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  round(prop.table(table(data[, "important.same.geographic.region"])) *     100, 0)[c(1, 3)]
## X-squared = 31.135, df = 1, p-value = 2.407e-08
```

```
round(prop.table(table(data[,"important.same.geographic.region"])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##            18            82
```

How important is it working with people from same geographic region across gender? No.

```
chisq.test(percentage_distribution(data[,c("important.same.geographic.region","gender")])[c(1,3),])
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(data[, c("important.same.geographic.region",     "gender")])[c(1, 3),
## X-squared = 0.00042971, df = 1, p-value = 0.9835
```

```
percentage_distribution(data[,c("important.same.geographic.region","gender")])
```

```
##                                 gender
## important.same.geographic.region Man Woman
##                     Important      13     9
##                     Neutral        24    41
##                     Not important  64    50
```

How important is it working with people from same geographic region across regions? Yes.

```
chisq.test(percentage_distribution(data[,c("important.same.geographic.region","regionl1")])[c(1,3),])
```

```
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(data[, c("important.same.geographic.region",     "regionl1")])[c(1, 3
## X-squared = 20.605, df = 3, p-value = 0.0001271
```

```
percentage_distribution(data[,c("important.same.geographic.region","regionl1")])
```

```
##                                 regionl1
## important.same.geographic.region Europe Asia Americas Africa
##                     Important          7   14       10     22
##                     Neutral           22   25       38     44
##                     Not important     71   61       52     33
```

```
round(prop.table(table(data[,c("important.same.geographic.region","regionl1")]))[c(1,3),],2)*100,0)
```

```
##                                 regionl1
## important.same.geographic.region Europe Asia Americas Africa
##                     Important          9   19       15     40
##                     Not important     91   81       85     60
```

### Challenge working with people who speak different language

Not conclusive.

```
chisq.test(round(prop.table(table(data[,"challenging.speak.different.language"]))*100,0)[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  round(prop.table(table(data[, "challenging.speak.different.language"])) *     100, 0)[c(1, 3)]
## X-squared = 0.51429, df = 1, p-value = 0.4733
```

```
round(prop.table(table(data[,"challenging.speak.different.language"]))*100,0)
```

```
##
##     Challenging         Neutral Not challenging
##              32              30              38
```

How challenging is it to work with people who speak different language across gender? No difference.

```
chisq.test(percentage_distribution(data[,c("challenging.speak.different.language","gender")])[c(1,3),])
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(data[, c("challenging.speak.different.language",     "gender")])[c(1,
## X-squared = 0.34614, df = 1, p-value = 0.5563
```

```
percentage_distribution(data[,c("challenging.speak.different.language","gender")])
```

```
##                                     gender
## challenging.speak.different.language Man Woman
##                      Challenging      33    26
##                      Neutral          29    35
##                      Not challenging  38    39
```

How challenging is working with people who speak different language across regions? Differences exist.

```
chisq.test(percentage_distribution(data[,c("challenging.speak.different.language","regionl1")])[c(1,3),]
```

```
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(data[, c("challenging.speak.different.language",     "regionl1")])[c(
## X-squared = 28.281, df = 3, p-value = 3.171e-06
```

```
percentage_distribution(data[,c("challenging.speak.different.language","regionl1")])
```

```
##                                     regionl1
## challenging.speak.different.language Europe Asia Americas Africa
##                      Challenging         16   41       43     33
```

```
##                       Neutral             40   17       14     58
##                       Not challenging     44   41       43      8
```

```
round(prop.table(table(data[,c("challenging.speak.different.language","regionl1")])[c(1,3),],2)*100,0)
```

```
##                                     regionl1
## challenging.speak.different.language Europe Asia Americas Africa
##                       Challenging         26   50       50     80
##                       Not challenging     74   50       50     20
```

### How helpful are translation tools?

Not conclusive

```
chisq.test(round(prop.table(table(data[,"helpful.translation.tools"]))*100,0))
```

```
##
##  Chi-squared test for given probabilities
##
## data:  round(prop.table(table(data[, "helpful.translation.tools"])) *     100, 0)
## X-squared = 3.3465, df = 2, p-value = 0.1876
```

```
round(prop.table(table(data[,"helpful.translation.tools"]))*100,0)
```

```
##
##     Helpful     Neutral Not helpful
##          38          38          25
```

How helpful are translation tools across gender? Women find it more useful than men.

```
chisq.test(percentage_distribution(data[,c("helpful.translation.tools","gender")])[c(1,3),])
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(data[, c("helpful.translation.tools",     "gender")])[c(1, 3), ]
## X-squared = 6.4939, df = 1, p-value = 0.01082
```

```
percentage_distribution(data[,c("helpful.translation.tools","gender")])
```

```
##                           gender
## helpful.translation.tools Man Woman
##                Helpful      31    57
##                Neutral      43    26
##                Not helpful  26    17
```

```
round(prop.table(table(data[,c("helpful.translation.tools","gender")])[c(1,3),],2)*100,0)
```

```
##                           gender
## helpful.translation.tools Man Woman
##                Helpful      55    76
##                Not helpful  45    24
```

How helpful are translation tools across regions? No difference

```
chisq.test(percentage_distribution(data[,c("helpful.translation.tools","regionl1")])[c(1,3),])
```

```
##
##  Pearson's Chi-squared test
```

```
##
## data:  percentage_distribution(data[, c("helpful.translation.tools",     "region1")])[c(1, 3), ]
## X-squared = 6.2523, df = 3, p-value = 0.09996
```

```
percentage_distribution(data[,c("helpful.translation.tools","region1")])
```

```
##                          region1
## helpful.translation.tools Europe Asia Americas Africa
##               Helpful         28   38       48     36
##               Neutral         41   45       24     45
##               Not helpful     30   17       29     18
```

### Importance of same gender identity

Not important

```
chisq.test(table(data[,c("important.same.gender.identity")])[c(1,3)])
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(data[, c("important.same.gender.identity")])[c(1, 3)]
## X-squared = 64.34, df = 1, p-value = 1.047e-15
```

```
table(data[,c("important.same.gender.identity")])
```

```
##
##     Important       Neutral Not important
##             9            20            88
```

```
round(prop.table(table(data[,c("important.same.gender.identity")])[c(1,3)])*100,0)
```

```
##
##     Important Not important
##             9            91
```

How important is same gender identity across genders? Less important for men

```
chisq.test(percentage_distribution(data[,c("important.same.gender.identity","gender")])[c(1,3),])
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  percentage_distribution(data[, c("important.same.gender.identity",     "gender")])[c(1, 3), ]
## X-squared = 31.185, df = 1, p-value = 2.346e-08
```

```
percentage_distribution(data[,c("important.same.gender.identity","gender")])
```

```
##                              gender
## important.same.gender.identity Man Woman
##                  Important       1    30
##                  Neutral        18    17
##                  Not important  81    52
```

```
round(prop.table(table(data[,c("important.same.gender.identity","gender")])[c(1,3),],2)*100,0)
```

```
##                              gender
## important.same.gender.identity Man Woman
##                  Important       1    37
```

```
##                   Not important  99    63
```

How important is same gender identity across regions? Different for different regions.

```
chisq.test(percentage_distribution(data[,c("important.same.gender.identity","regionl1")])[c(1,3),])
```

```
##
##  Pearson's Chi-squared test
##
## data:  percentage_distribution(data[, c("important.same.gender.identity",     "regionl1")])[c(1, 3),
## X-squared = 17.546, df = 3, p-value = 0.0005457
```

```
percentage_distribution(data[,c("important.same.gender.identity","regionl1")])
```

```
##                                regionl1
## important.same.gender.identity Europe Asia Americas Africa
##                   Important          9    3       14      0
##                   Neutral           17   17       14     27
##                   Not important     74   79       71     73
```

```
round(prop.table(table(data[,c("important.same.gender.identity","regionl1")])[c(1,3),],2)*100,0)
```

```
##                                regionl1
## important.same.gender.identity Europe Asia Americas Africa
##                   Important         11    4       17      0
##                   Not important     89   96       83    100
```