Mathematics Developers Survey

Nejc Ilenic

Where in the world and in which domain should a data scientist start his / her career?

Annual Developer Survey

Annual Developer Survey

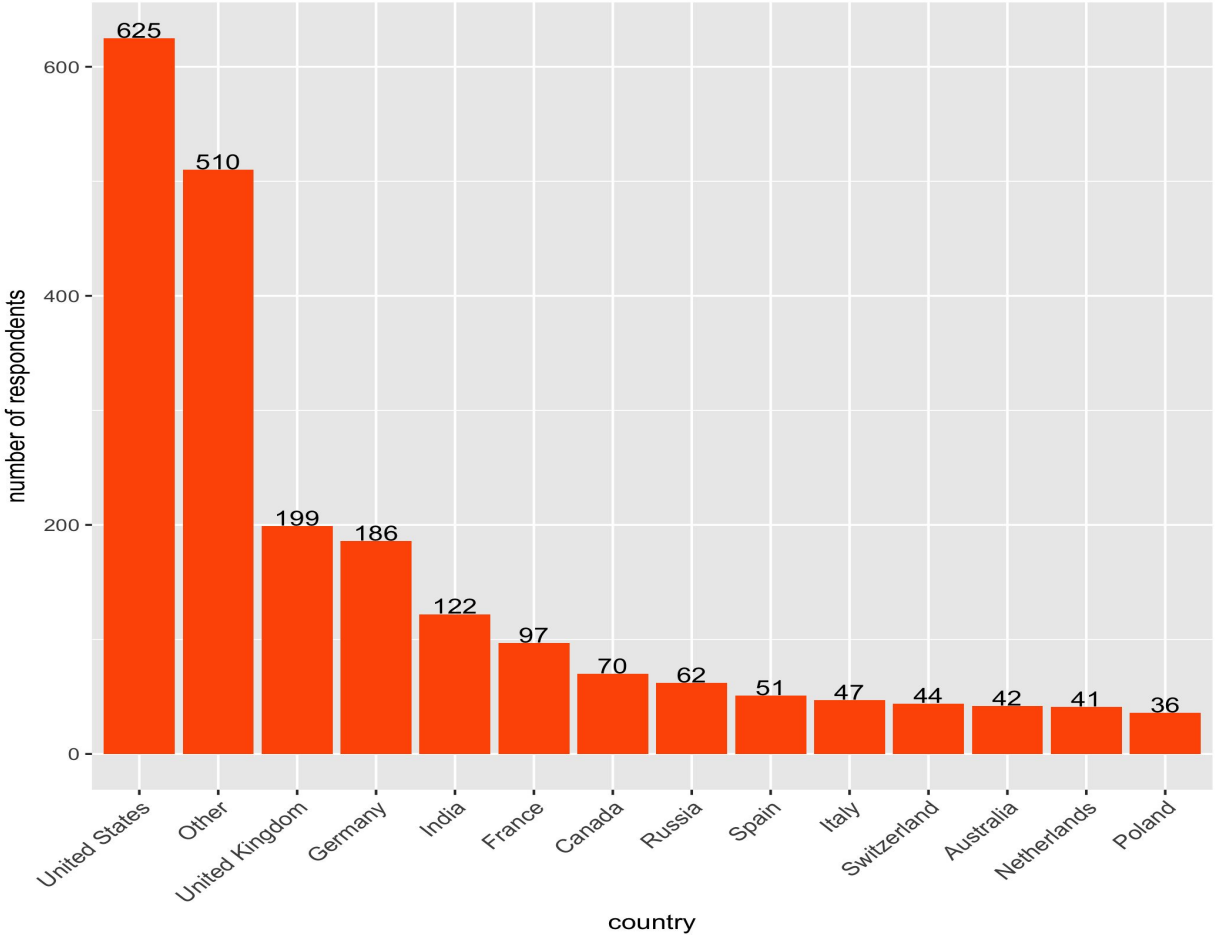
from January 7 to January 25, 2016

filter occupation group of mathematics developers

Annual Developer Survey

dataset **may not be a representative sample** from the population

Number of respondents per country



Job satisfaction comparison

’How satisfied are you with your current job(s)?’

Job satisfaction comparison

’I hate my job’
’I’m somewhat dissatisfied with my job’
’I’m neither satisfied nor dissatisfied’
’I’m somewhat satisfied with my job’
’I love my job’

Multinomial-Dirichlet model

$$n \text{ trials, } k \text{ categories}$$

$$y_i \in \{0, ..., n\}, \sum y_i = n$$

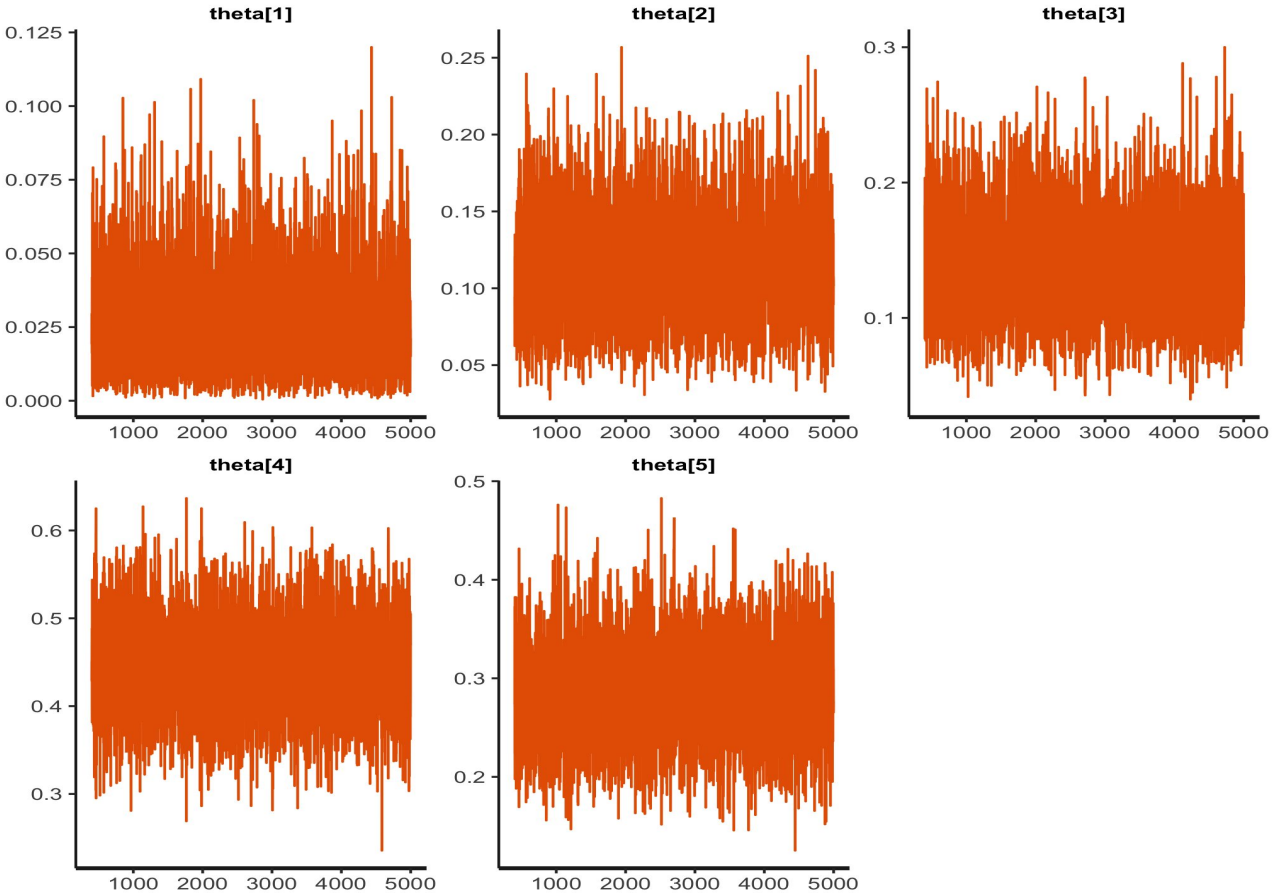
$$\theta \in k\text{-simplex}$$

$$y|\theta \underset{iid}{\sim} \text{multinomial}(\theta)$$

$$\theta \sim \text{Dirichlet}(\alpha)$$

$$\theta|y \sim \text{Dirichlet}(\alpha + y)$$

MCMC diagnostics (France)

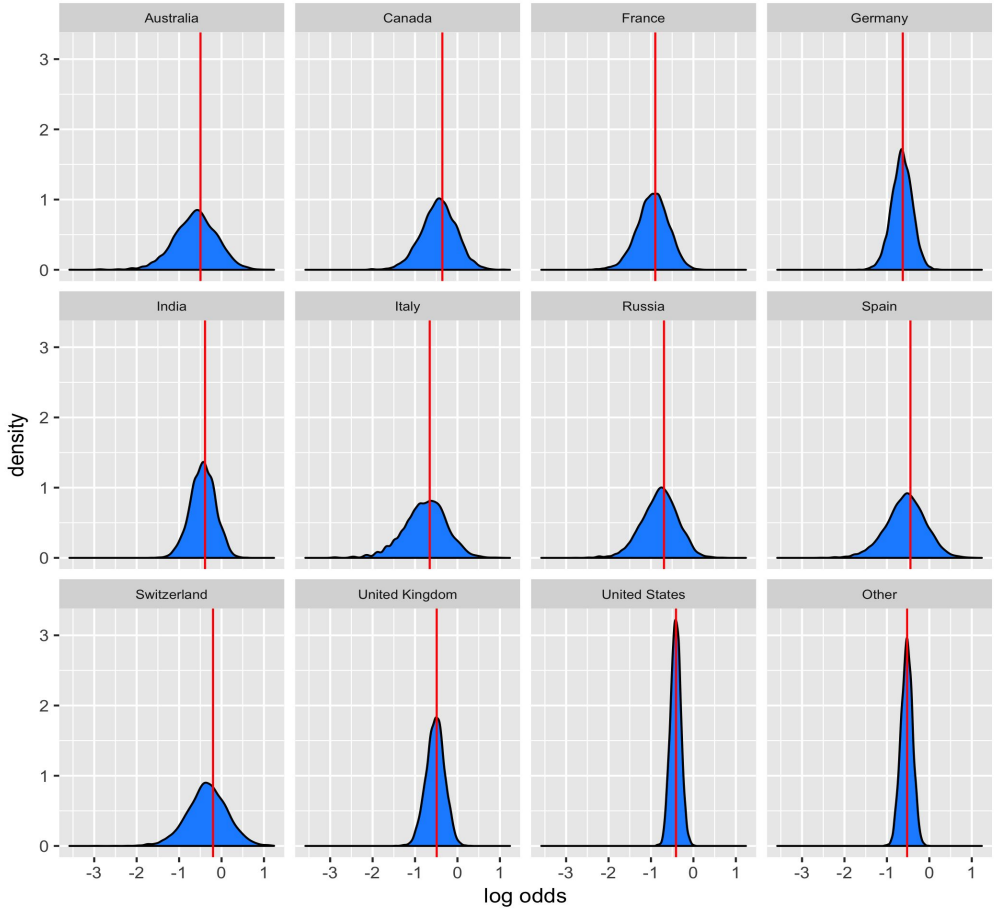


5000 iterations
400 warmup

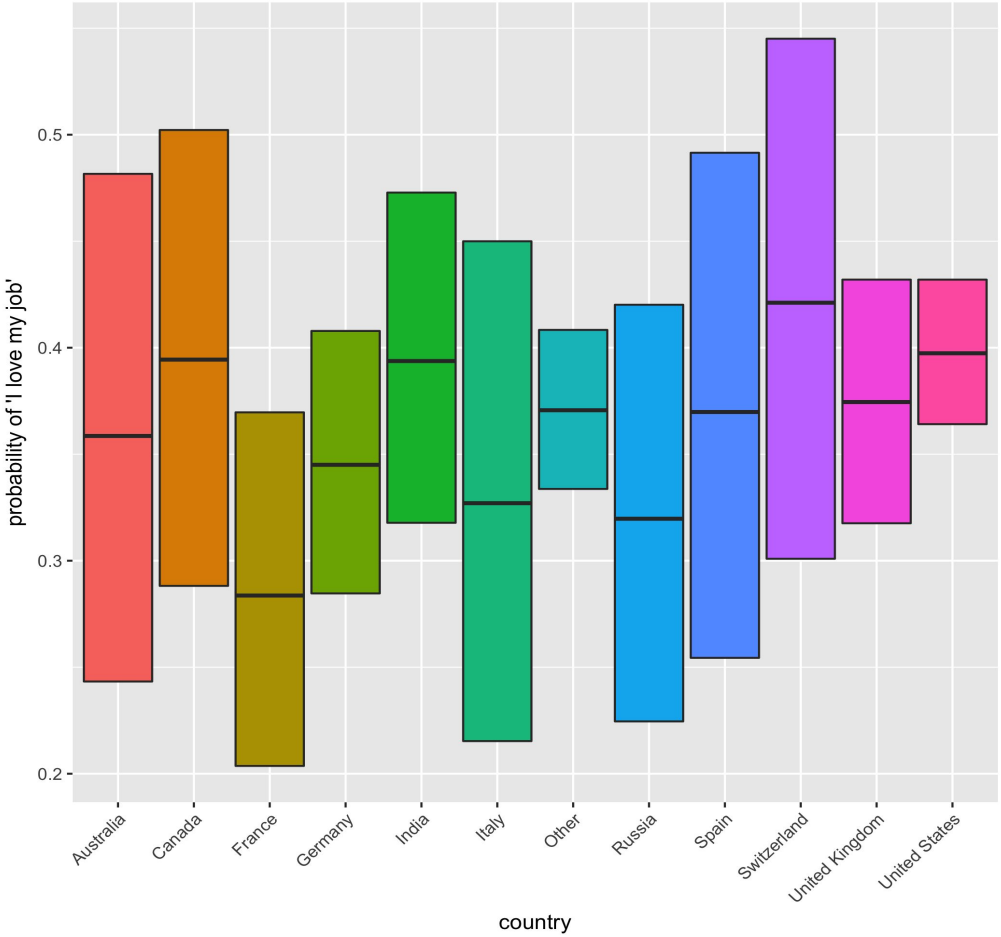
se_mcmc:
theta[1]: 0.00024
theta[2]: 0.00051
theta[3]: 0.00057
theta[4]: 0.00080
theta[5]: 0.00074

n_eff:
4600

Posterior predictive check: log odds of ’I love my job’



90% CIs for posterior probabilities of ’I love my job’



How job satisfaction depends on other factors?

outcome as an ordinal variable

model: **ordinal logistic regression**

Explanatory variables

age: continuous

gender: dummy

purchasing power: continuous

Purchasing power

compensation [$] / Big Mac index

i.e. how many Big Macs per year

Explanatory variables

works remotely: dummy

values unit testing: dummy

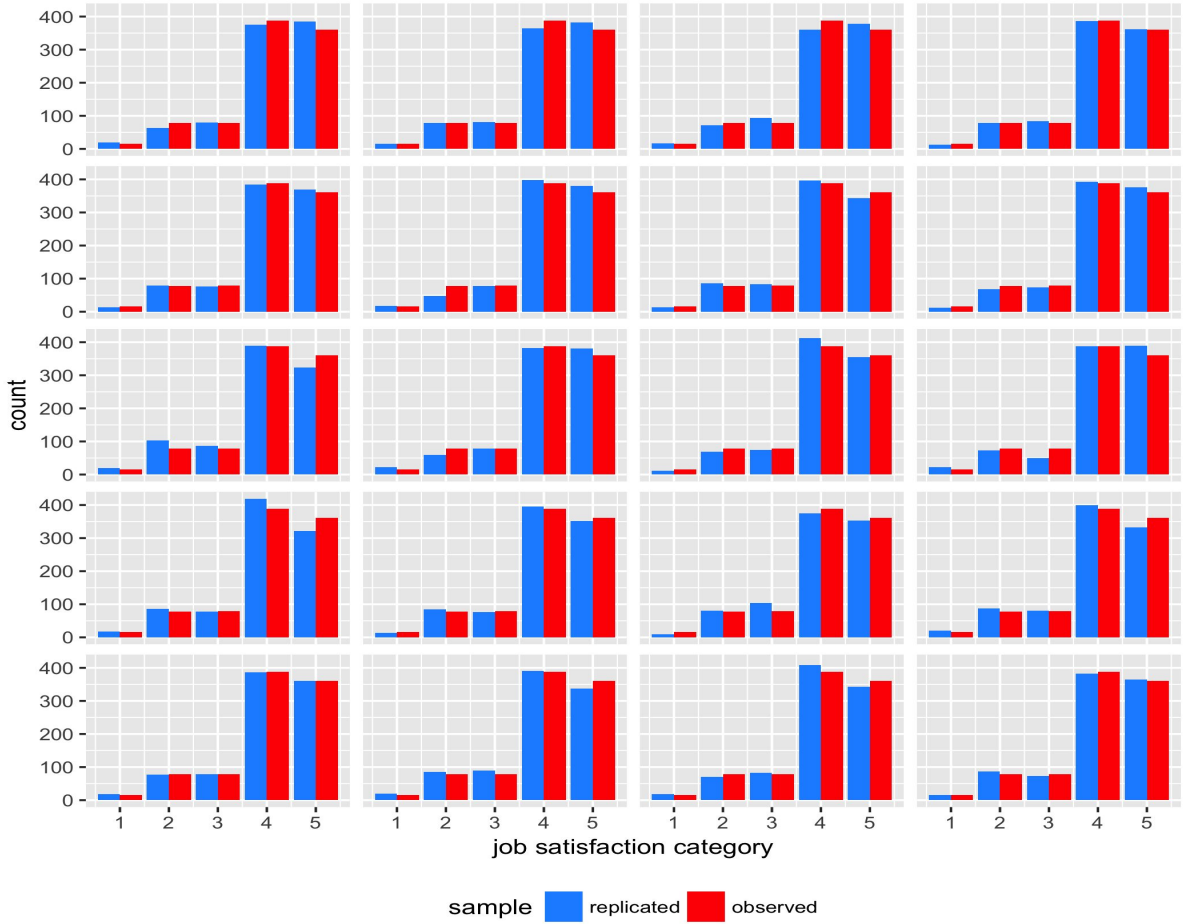
commits code once a day: dummy

Explanatory variables

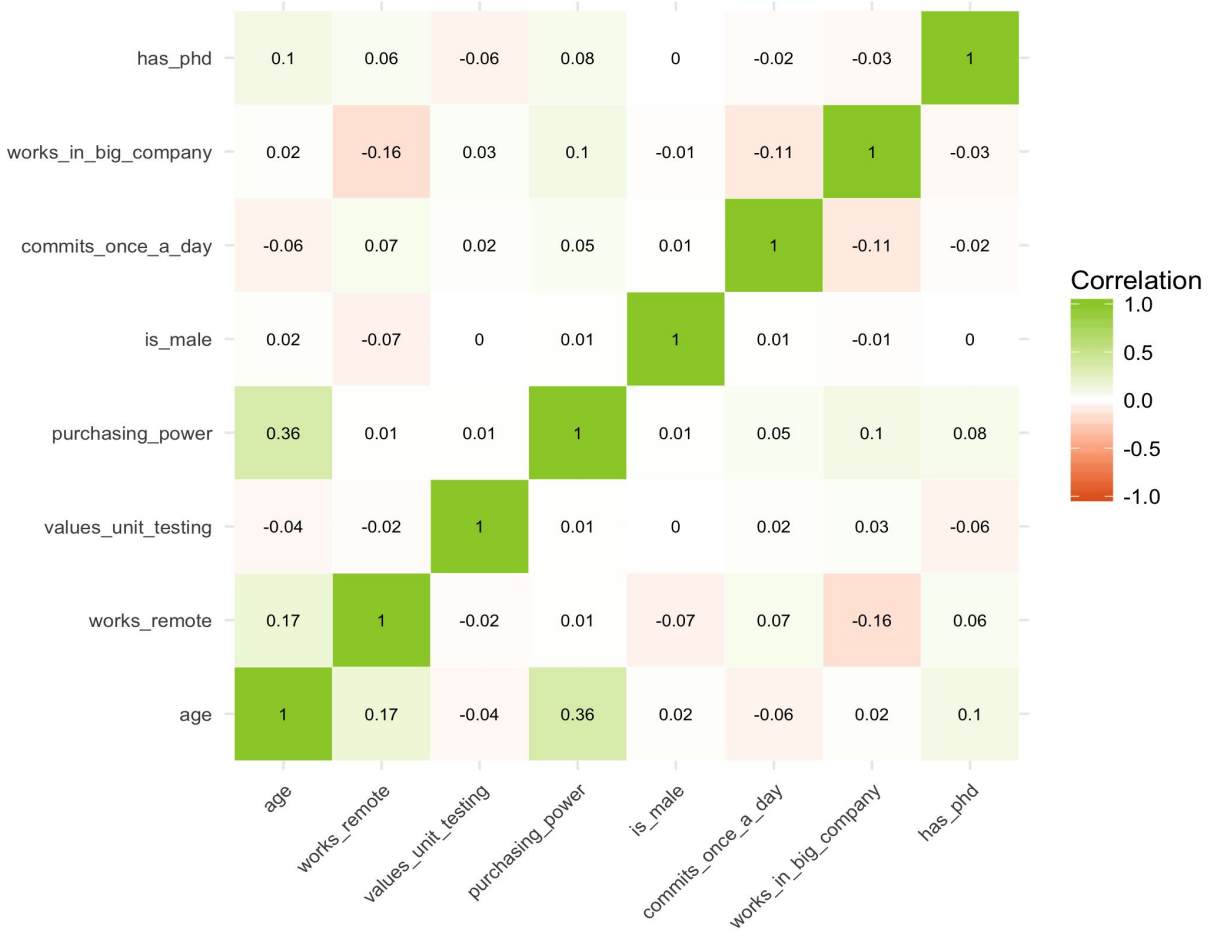
works in a big company (>99 employees): dummy

has PhD: dummy

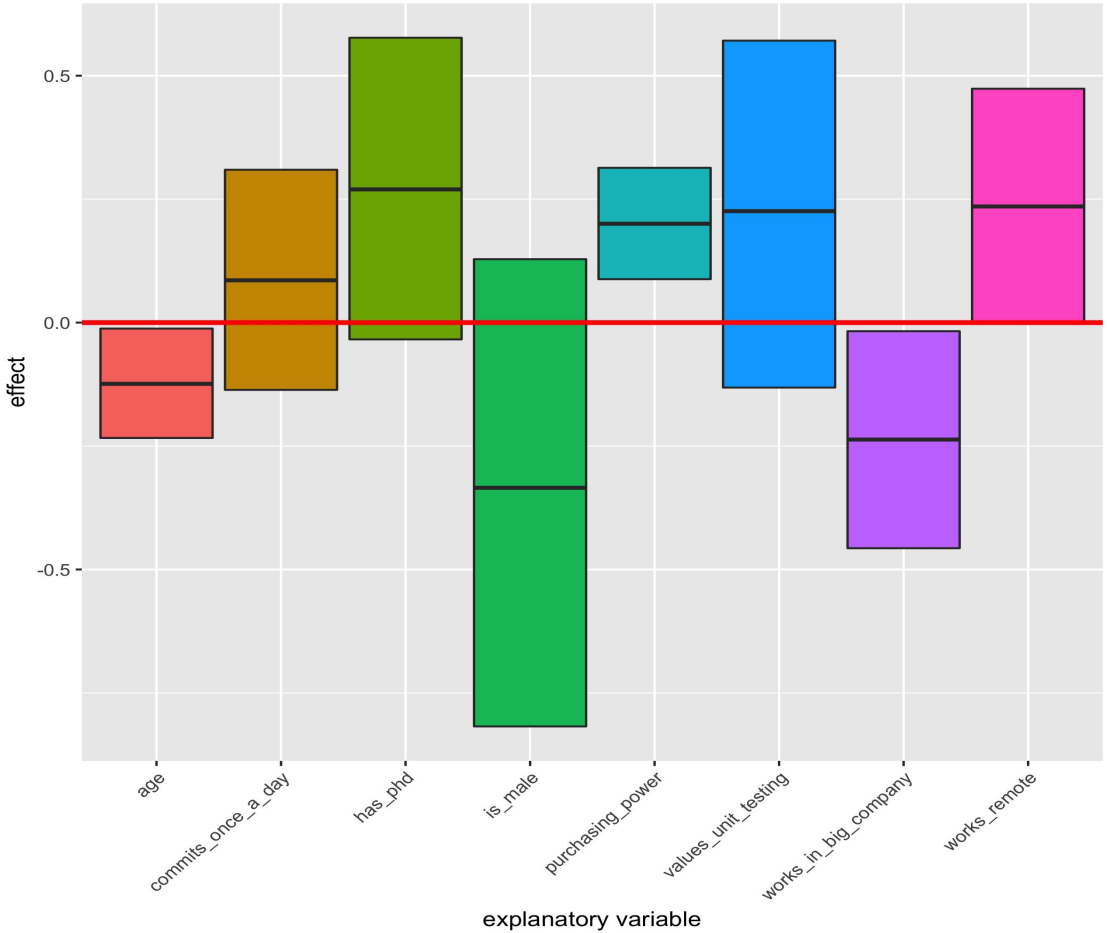
Posterior predictive check



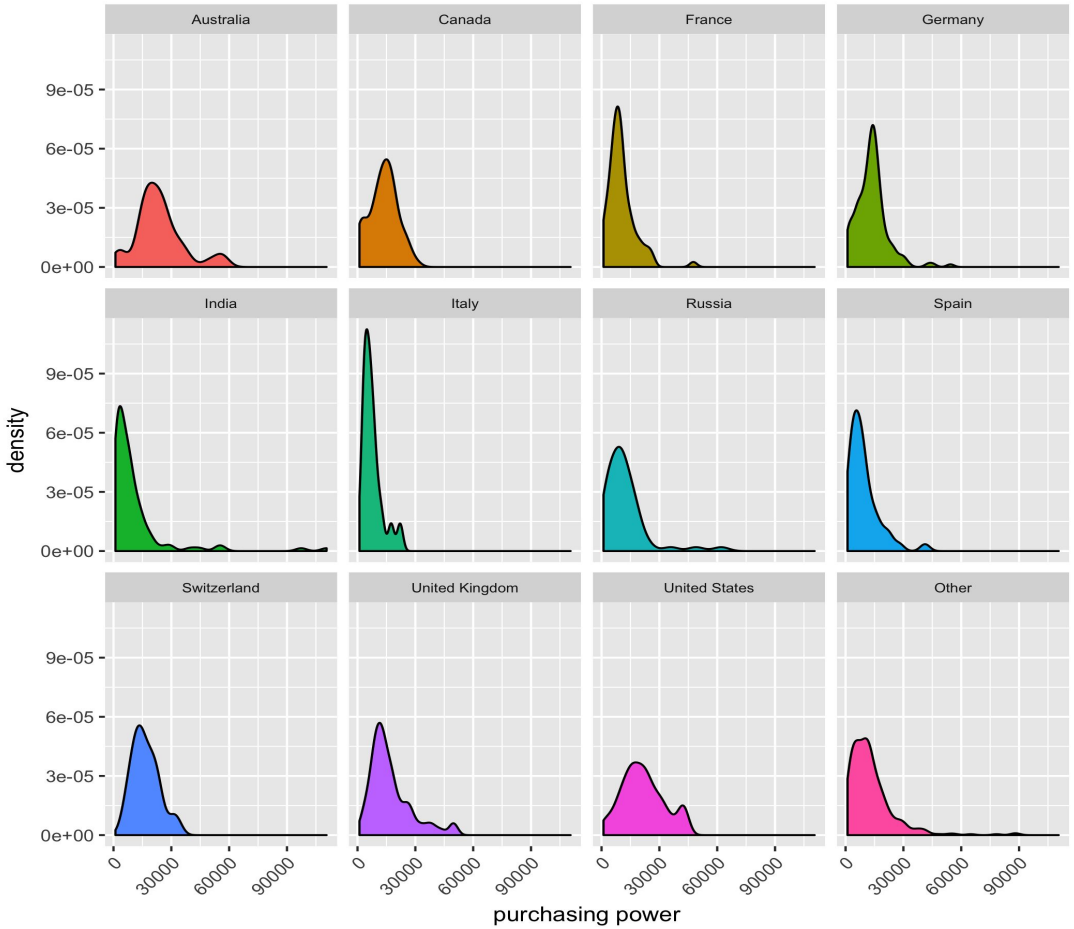
Explanatory variables: pearson correlation coefficients



90% CIs for posterior regression coefficients



Purchasing power comparison: observed densities

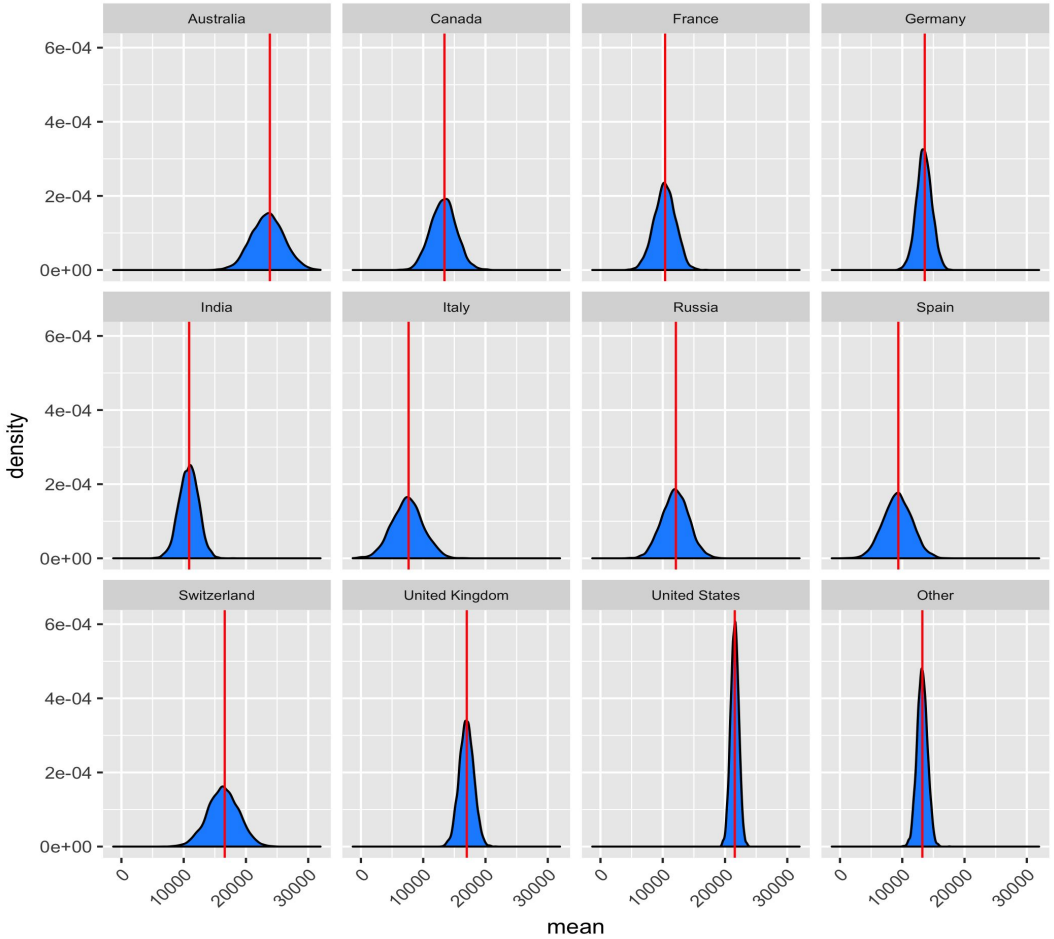


Purchasing power comparison

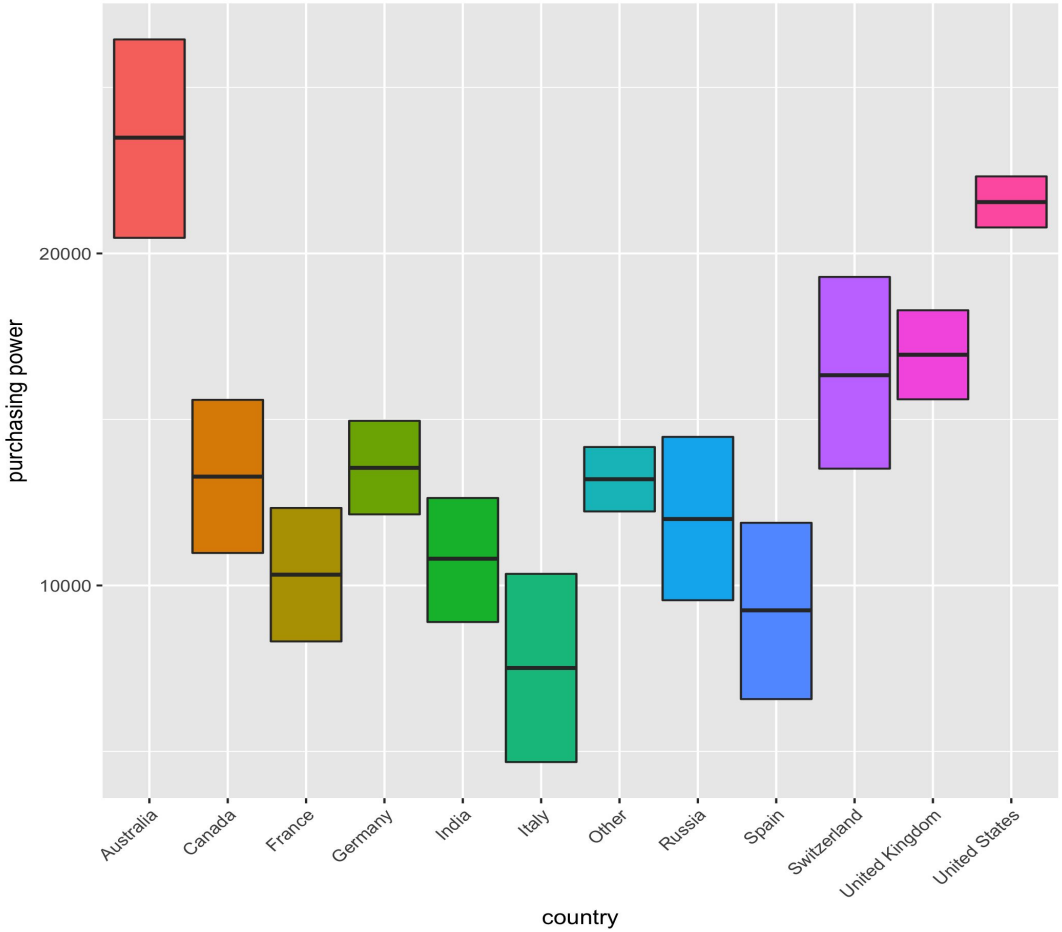
hierarchical normal model

within group variability assumed constant across groups

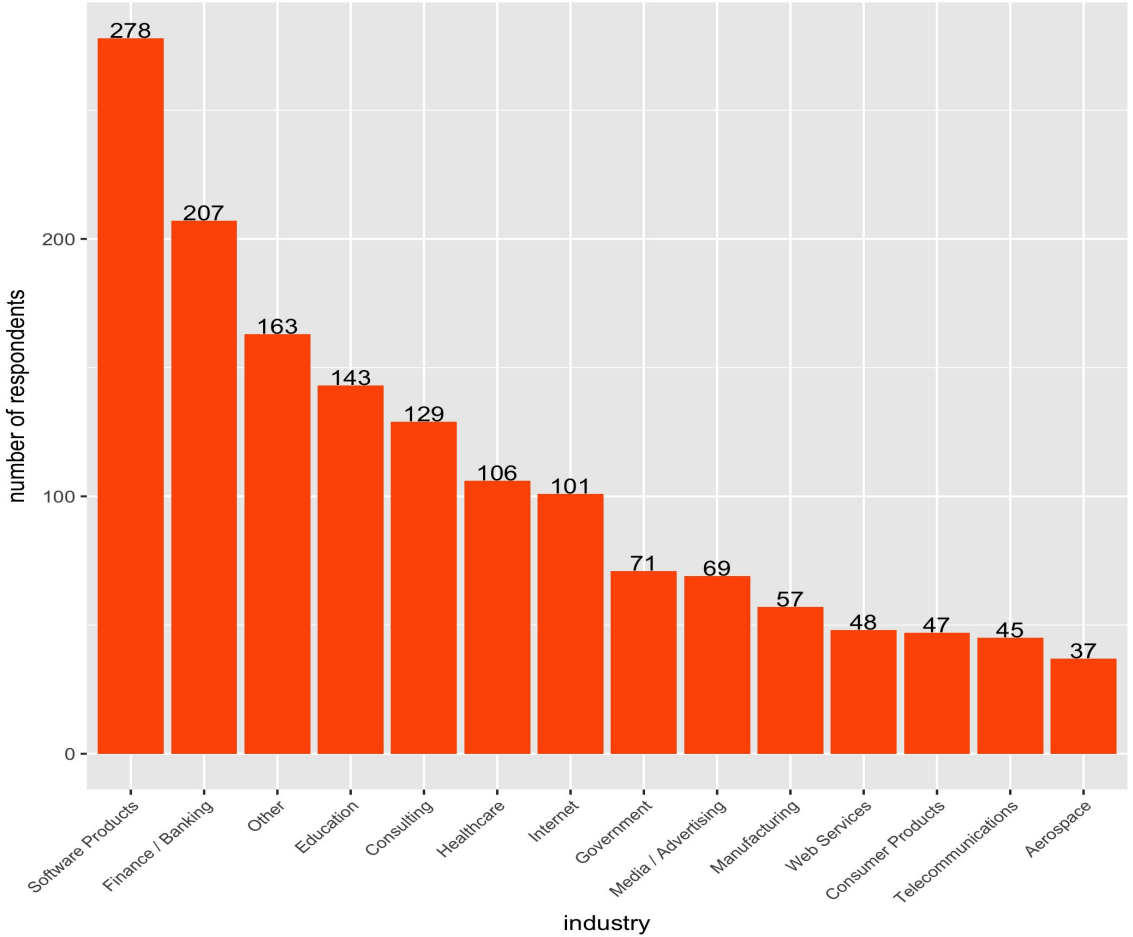
Posterior predictive check: means



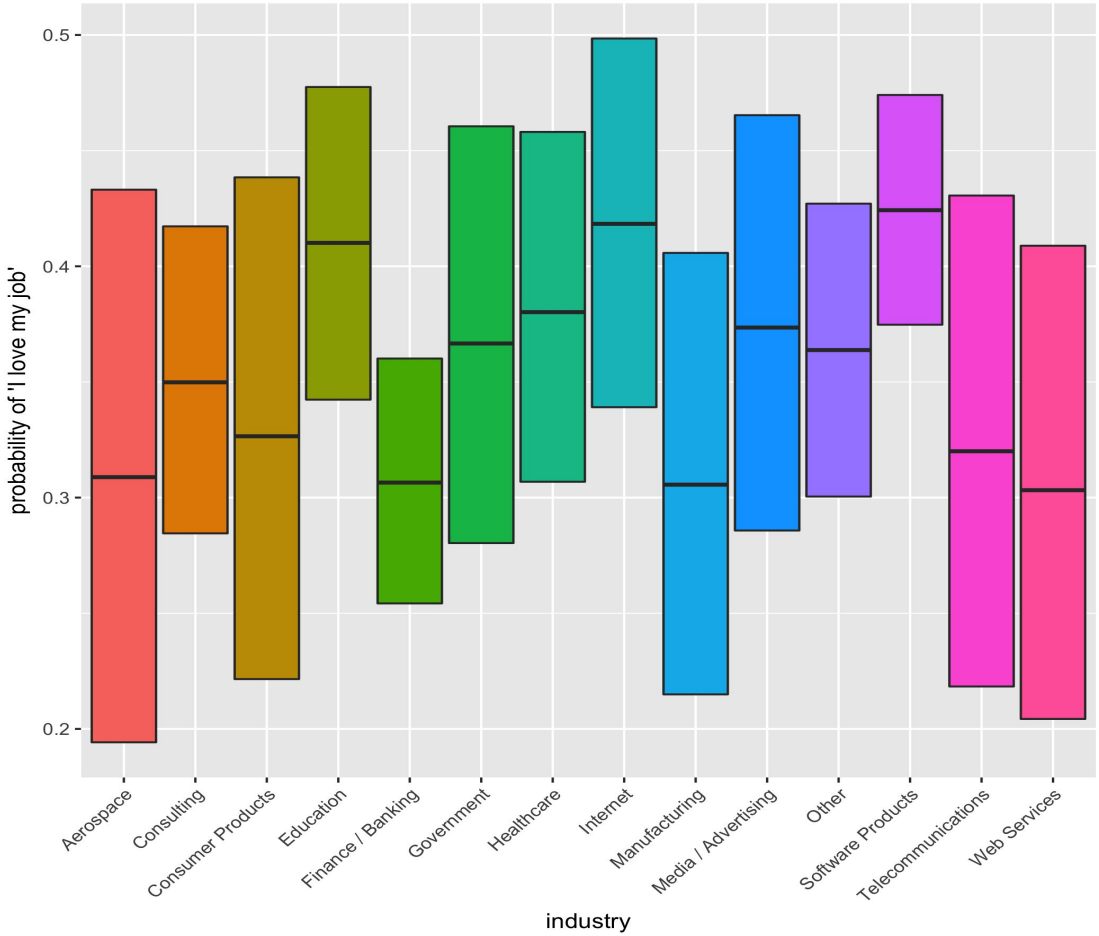
90% CIs for purchasing power posterior means



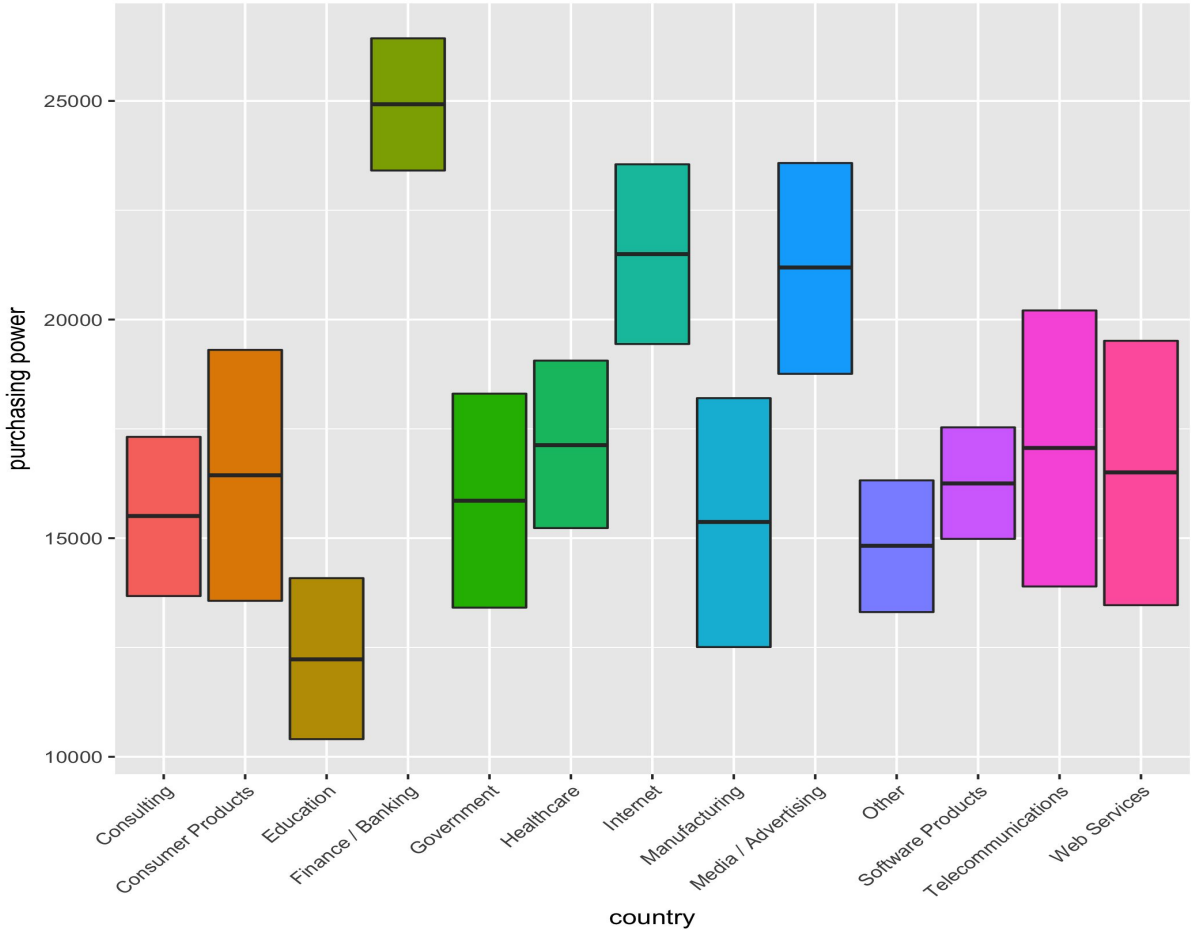
Number of respondents per industry



90% CIs for posterior probabilities of ’I love my job’



90% CIs for purchasing power posterior means



Recap

age and **working in a big company**
negatively correlated with job satisfaction

Recap

suspect **working remotely** positively correlated with job satisfaction

Recap

quite certain **purchasing power**
positively correlated with job satisfaction

Recap

mean **purchasing power** highest in Australia and United States

Recap

mean **purchasing power** highest in Finance / Banking (much higher than in education)