Pair Assignment 3

Philip & Philipp

14 Apr 2016

To do

- Rewrite

- Comment on last figure?

- Insert Bertrand and possibly other references

- Clean repository; put datasets in a folder.

- Re-write readme

- British or American English?

Motivation

The research question we wish to pursue for the final project is: *"Are career pursuits reconcilable with a happy life?"*. Our hypothesis is that under some conditions, pursuing a career makes it difficult to reconcile high work intensity with a family or a social life, which could decrease overall life-satisfaction. The link could be particularly pronounced for women. First, gender norms may impose a happiness penalty, if pursuing a career clashes with dominent gender roles in the society. Second, when both men and women adhere to gender norms and follow a single bread-winner model they may benefit additionally to the norm reward from fewer tensions between career pursuits and family life as genders each specialize on one domain (Reference).

To analyse the question we are using three data-sources. The General Social Survey from the United States, which is ongoing since 1972 and ask survey respondents questions regarding well-being in different dimensions (overall, work, family), demographic characteristics, job-affiliation, income and more. Further, we use the Personal Consumption Expenditure index from the Federal Reserve Bank of St. Louis to adjust income variables for inflation. Last, we use data from the Current Population Survey on income percentiles in age and educational groups, which is collected from the replication file of Bertrand (2013). [Footnote: Bertrand only includes the income percentiles between 1977-2010, and at a later stage we intend to calculate the income percentiles directly from the CPS].

1

Based on the available data-sources there are considerable methodological challenges, of which we will elaborate on the most pronounced. First, defining the group of individuals who pursue a career is problematic. In the literature, career pursuits have been proxied by income thresholds (Goldin, 2004; Bertrand, 2013). However, these measures neglect that individuals can have career ambitions in low-income jobs, and not all high earners are pursuing a career. In this paper we also deploy an income-threshold variable, but interpret it simply as indicating high-income workers. Second, when analysing sub-populations of the General Social Survey, we face issues of small sample sizes. The GSS has surveyed around 60.000 individuals between 1972 and 2014, but when comparing individuals with specific educational attainment, labour-force participation, family status, gender etc., the sample size can be reduced to a few hundred observations. Third, there is a large literature on whether it is possible to draw inference on subjective well-being measures, see for instance Bertrand (2013), xx or xx.

The scope of this paper is to combine data from the three above mentioned sources and present descriptive statistics and correlations that are of importance with respect to our research question. All steps are done dynamically, such that they are easily reproducible.

Descriptive results

Distribution of overall and life satisfaction

Two of the central variables for our research is reported happiness and job-satisfaction, which are based on survey respondents' answers to the two questions: "Taken all together, how would you say things are these days?" and "On the whole, how satisfied are you with the work you do?". Figure 1 shows the distribution of answers to the two questions. The former is measured on a three point scale (higher is better), of which around 60 % of the sample respond a middle happiness level. The latter is measured on a four point scale, and figure 1 shows that a majority of the sample responds either moderate or high job satisfaction (score 3 and 4).

Figure 1: Distribution of Life- and Work-Satisfaction

```
happy_dist <- table(z$happy)/sum(table(z$happy))*100
satjob_dist <- table(z$satjob)/sum(table(z$happy))*100


par(mfrow=c(1,2))
```

```
barplot(happy_dist,

        col=c("navyblue"),

        xlab = "Overall life satisfaction categories",

        legend = c("Life satisfaction"),

        ylim = c(0, 80))


barplot(satjob_dist,

        col=c("navyblue"),

        xlab = "Job satisfaction categories",

        legend = c("Job satisfcation"),

        ylim = c(0, 80))
```


Overall life satisfaction categories   Job satisfaction categories

Average happiness in different survey years

The GSS is conducted between 1972 and 2014. Due to year specific events, unintended differences in the implementation of the survey or trends in overall happiness, there can be year-specific differences. Figure 2 shows the average share of the population who report high-happiness across the survey years.

Figure 2: Reported Life-Satisfaction over Time

The figure shows that there is considerable variation between years, and a weak negative trend particularly between 1972 and 1983. It is not directly possible to disentangle what can be attributed to random noise and what is caused by structural changes, however, it signifies that it is pragmatic to control for survey year in a regression design to control for survey year fixed effects.

Happiness across age

Figure 3 investigates the relationship between reported happiness and age for college educated men and women.

```
ggplot() +
  stat_summary(data = z[z$sex == 1 & z$educat == 4,], aes(x=factor(age), y=vhappy),
               fun.y="mean", geom="point", col="Navyblue") +
  stat_summary(data = z[z$sex == 2 & z$educat == 4,], aes(x=factor(age), y=vhappy),
               fun.y="mean", geom="point", col="Red") +
  scale_x_discrete("Age") +
  scale_y_continuous("Percentage very happy") +
  expand_limits(y=c(20,60)) +
  theme_bw()
```
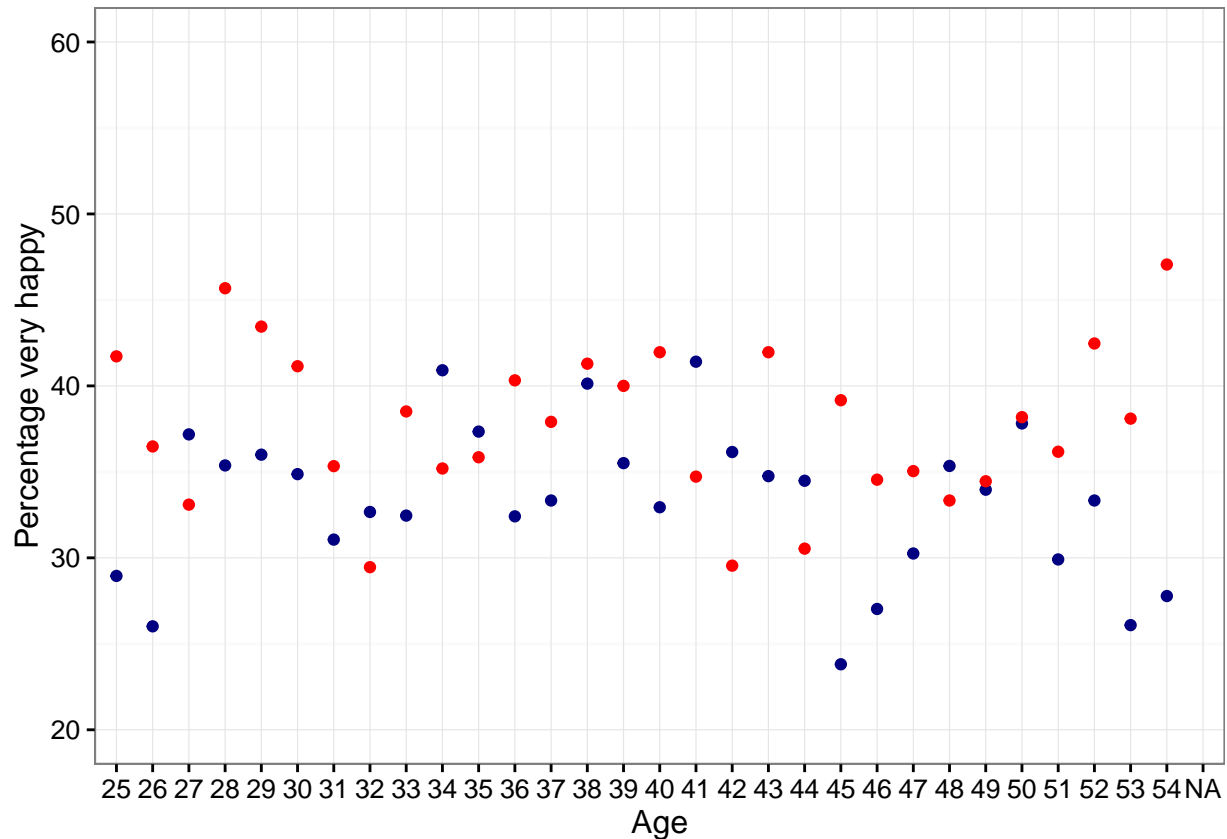
Figure 3: Happiness across Age (college educated men and women)

In the GSS there is no apparent structural relationship between the share of respondents who report being very happy and age. Further, college educated women have a slightly higher average reported happiness level relative to men (38% vs. 34%).

Distribution of work-hours for full-time workers

Work intensity is another variable of interest for the final research design. "Hours worked last week" is available across the full survey from 1972-2014. This is an imperfect proxy for respondents average working hours over a full year, but this unfortunately is not available. Figure x shows the frequency of hours worked last week for full-time workers. It is apparent that most respondents have worked around 40 hours a week, but also that there is considerable variance.

```
par(mfrow=c(1,1))
satjob_dist <- table(z$satjob)/sum(table(z$happy))*100
hist(z$hrs1[z$working_ft==1],
```

```
    col=c("navyblue"),

    border = "White",

    xlab = "Hours worked last week (full-time employees)",

    main = ""
)
```
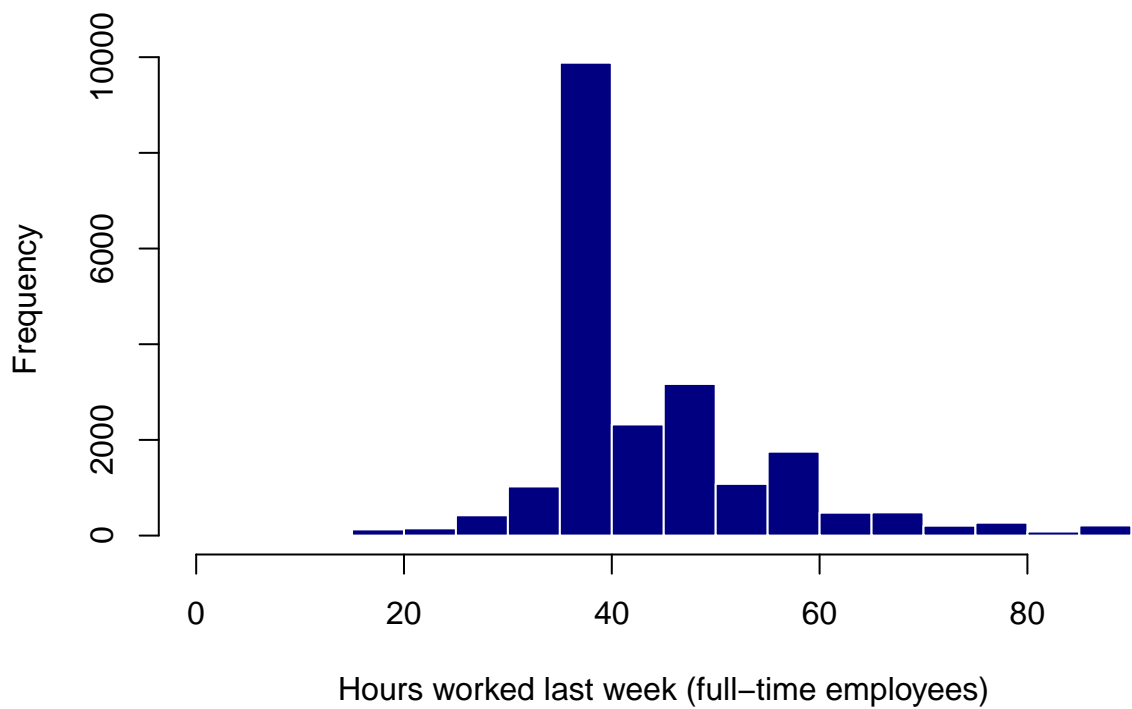

Gender

In today's society the strive for gender equality often clashes with labour market realities and the persistence of traditional gender norms. Previous studies found that the reconcilability of having a family and pursuing a career is still difficult to achieve for women, while it is rather the norm for men (e.g. Hipp reference). Thus, women who clash with traditional customs might face a "life happiness penalty" as shown in Bertrand (2013).

Figure x shows how reported happiness depends on labour-market affiliation for men (blue) and women (red) with a college degree. It shows that men report are substantially more likely to report being very happy when in full-time employment relative to part-time employment, which is not the case for women. Further, both men and women report high happiness levels when keeping house. Note, in the full sample only 35 college educated men are keeping house (there are 650 women) and when looking at all men, the average happiness score for men keeping house is only 24 %.

Figure 3: Happiness levels for college educated men and women depending on labour-market affiliation

```r
z$workstatus <- NA
z$workstatus[z$wrkstat==1] <- "Full-time work"
z$workstatus[z$wrkstat==2] <- "Part-time work"
z$workstatus[z$wrkstat==7] <- "Keeping house"



ggplot() +
  stat_summary(data = z[z$sex == 2 & z$educat==4,], aes(x=factor(workstatus), y=vhappy),
               fun.y="mean", geom="point", col="Red") +
  stat_summary(data = z[z$sex == 1 & z$educat ==4,], aes(x=factor(workstatus), y=vhappy),
               fun.y="mean", geom="point", col="Navyblue") +
  expand_limits(y=c(20,60)) +
  scale_x_discrete("Work status", limits=c("Full-time work", "Part-time work",
                "Keeping house")) +
  scale_y_continuous("Percentage very happy") +
  theme_bw()
```
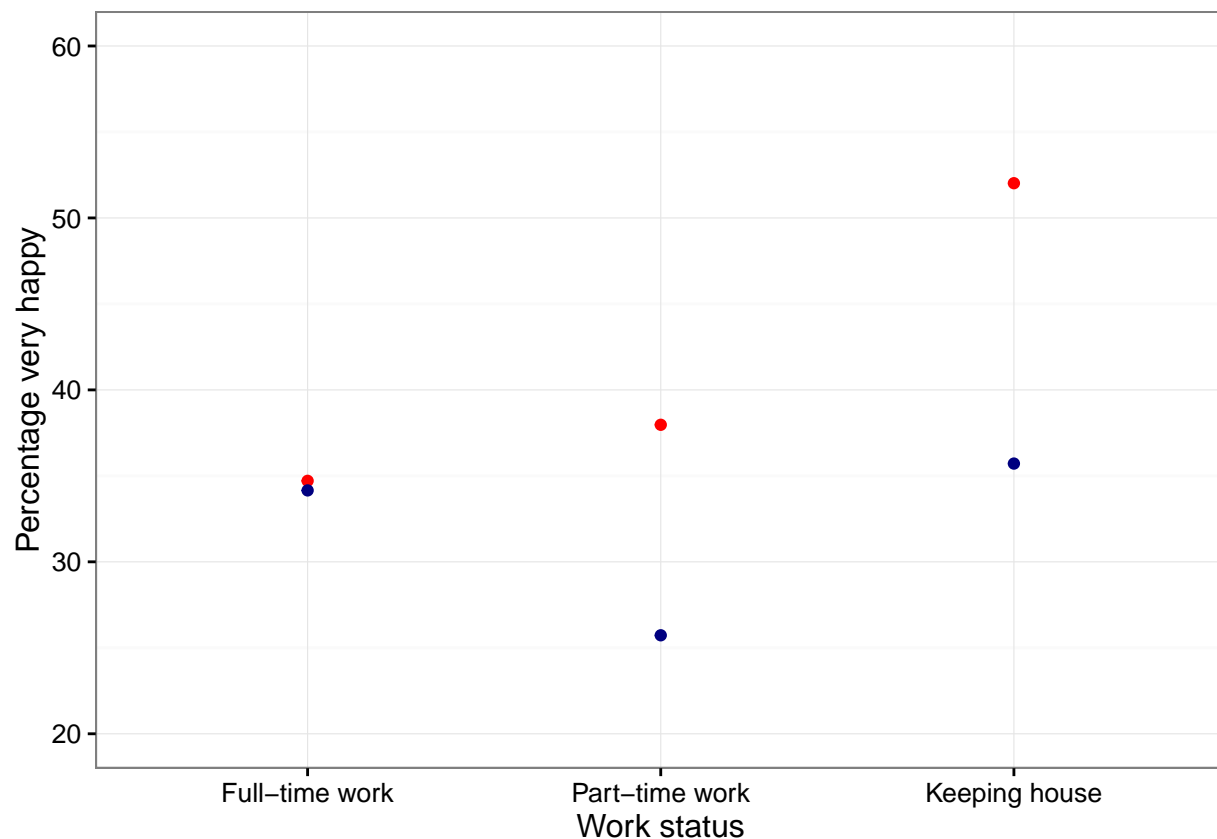
Figure X shows the share of college educated men and women who report being very happy depending on whether they earn more than the 25th (*or 50th*) income percentile of college educated men in their age cohort. The graph suggests that gender only plays a role when people do not have a high income. For women (red) the probability to be very happy does not change with having a low income. Men (blue) on the other hand face a large happiness reduction when having a low income. This penalty is even more pronounced when setting the threshold at the 50th percentile where high-income men report even higher hapiness than women.

Figure X: Happiness for women and men depending income status

```r
z = z[z$year >= 1977 & z$year < 2012,]


z$income_status <- NA
z$income_status[z$career==1] <- "High-income"
z$income_status[z$career==0] <- "Low-income"


z$career1 <- z$rinc > z$p50
z$career1 <- as.numeric(z$career1)
```
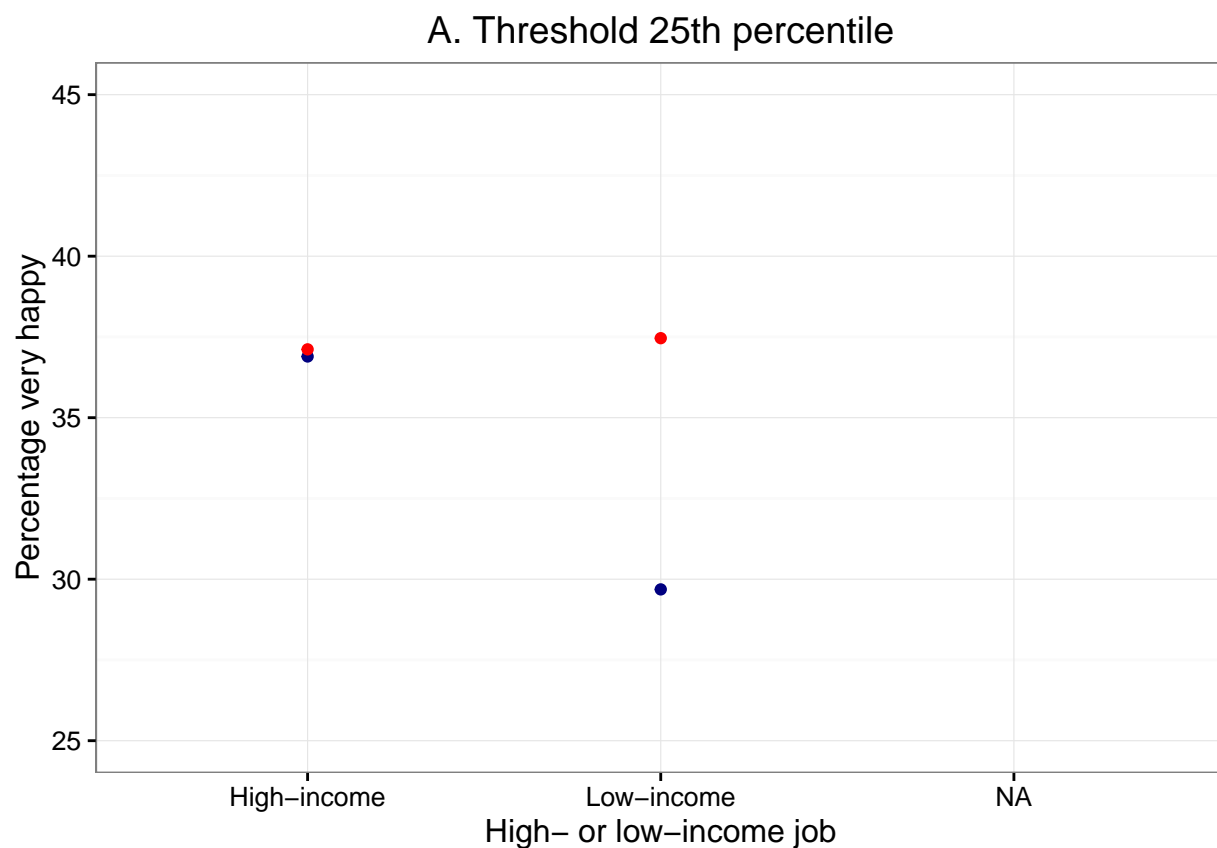
```
z$career1[is.na(z$career1)] <- 0


ggplot() +
  stat_summary(data = z[z$sex == 1 & z$educat==4,], aes(x=factor(income_status), y=vhappy),
               fun.y="mean", geom="point", col="Navyblue") +
  stat_summary(data = z[z$sex == 2  & z$educat==4,], aes(x=factor(income_status), y=vhappy),
               fun.y="mean", geom="point", col="Red") +
  expand_limits(y=c(25,45)) +
  scale_x_discrete("High- or low-income job") +
  scale_y_continuous("Percentage very happy") +
  ggtitle("A. Threshold 25th percentile") +
  theme_bw()
```

A. Threshold 25th percentile

```r
z$career1 <- z$rinc > z$p50

z$career1 <- as.numeric(z$career1)

z$career1[is.na(z$career1)] <- 0


z$income_status1 <- NA

z$income_status1[z$career1==1] <- "High-income"

z$income_status1[z$career1==0] <- "Low-income"


ggplot() +
  stat_summary(data = z[z$sex == 1 & z$educat==4,], aes(x=factor(income_status1), y=vhappy),
               fun.y="mean", geom="point", col="Navyblue") +
  stat_summary(data = z[z$sex == 2  & z$educat==4,], aes(x=factor(income_status1), y=vhappy),
               fun.y="mean", geom="point", col="Red") +
  expand_limits(y=c(25,45)) +
  scale_x_discrete("High- or low-income job") +
  scale_y_continuous("Percentage very happy") +
  ggtitle("B. Threshold 50th percentile") +
  theme_bw()
```
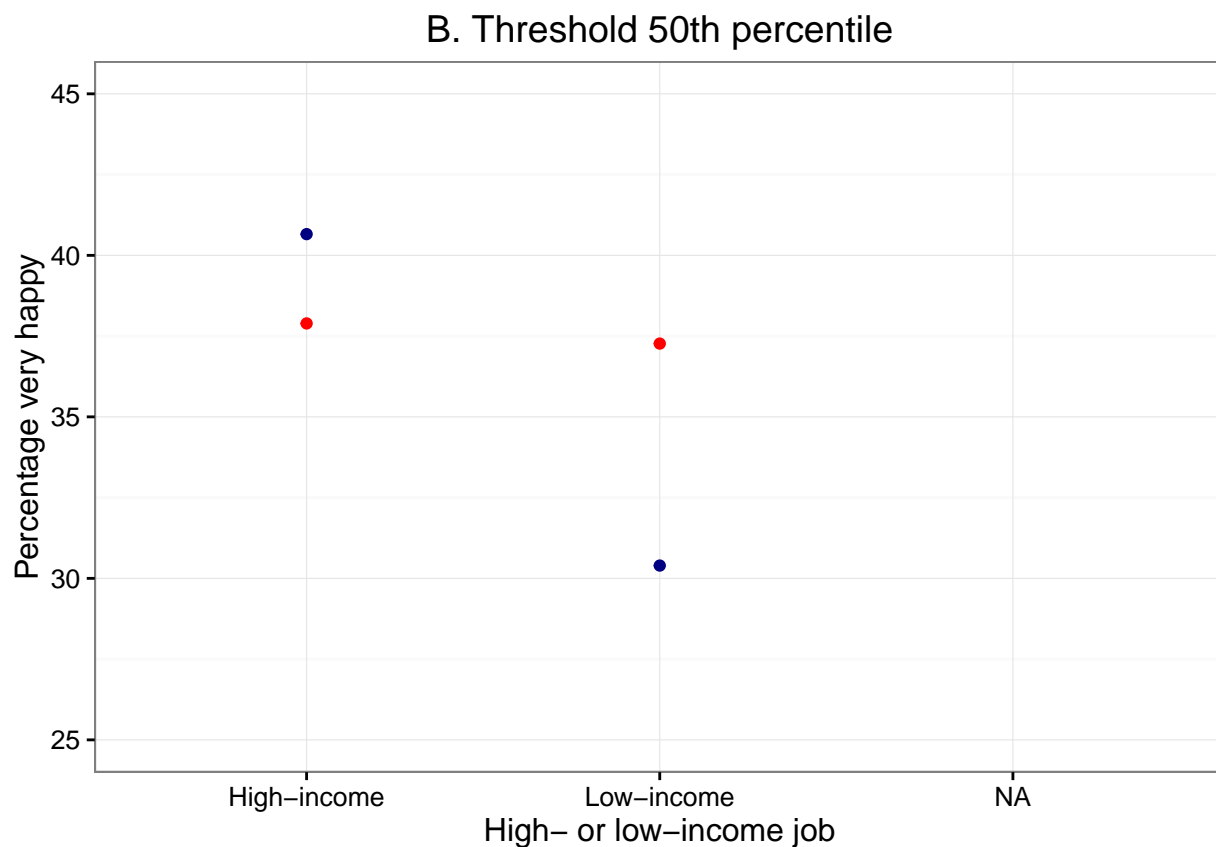
B. Threshold 50th percentile

Figure X further differentiates between four possible combinations of having a family (married and children) and having a high income job. Both college educated men and women report substantially higher happiness when having a family. When not having a family, higher income improves life satisfaction for both genders although the increase is slightly larger for men. Differences become more pronounced when having a family. With a family but without a high-income job women are the happiest, while men are considerably less happy, compared to having a career and a family. Thus, men and women interestingly have different career-family configurations under which they appear most happy.

Figure X: Reported happiness for college educated women and men depending on combination of having a high-income job and a family

```
z$meanhap <- NA
z$meanhap[z$family==0 & z$career==0] <- "No career, no family"
z$meanhap[z$family==0 & z$career==1] <- "Career, no family"
z$meanhap[z$family==1 & z$career==0] <- "No career, family"
z$meanhap[z$family==1 & z$career==1] <- "Career, family"
```

```
ggplot() +

  stat_summary(data = z[z$sex == 1 & z$educat==4 ,], aes(x=factor(meanhap), y=vhappy),

               fun.y="mean", geom="point", col="Navyblue") +

  stat_summary(data = z[z$sex == 2 & z$educat==4,], aes(x=factor(meanhap), y=vhappy),

               fun.y="mean", geom="point", col="Red") +

  scale_x_discrete("Family-work combination", limits=c("No career, no family",

               "Career, no family", "No career, family", "Career, family")) +

  scale_y_continuous("Percentage very happy") +

  theme_bw()
```
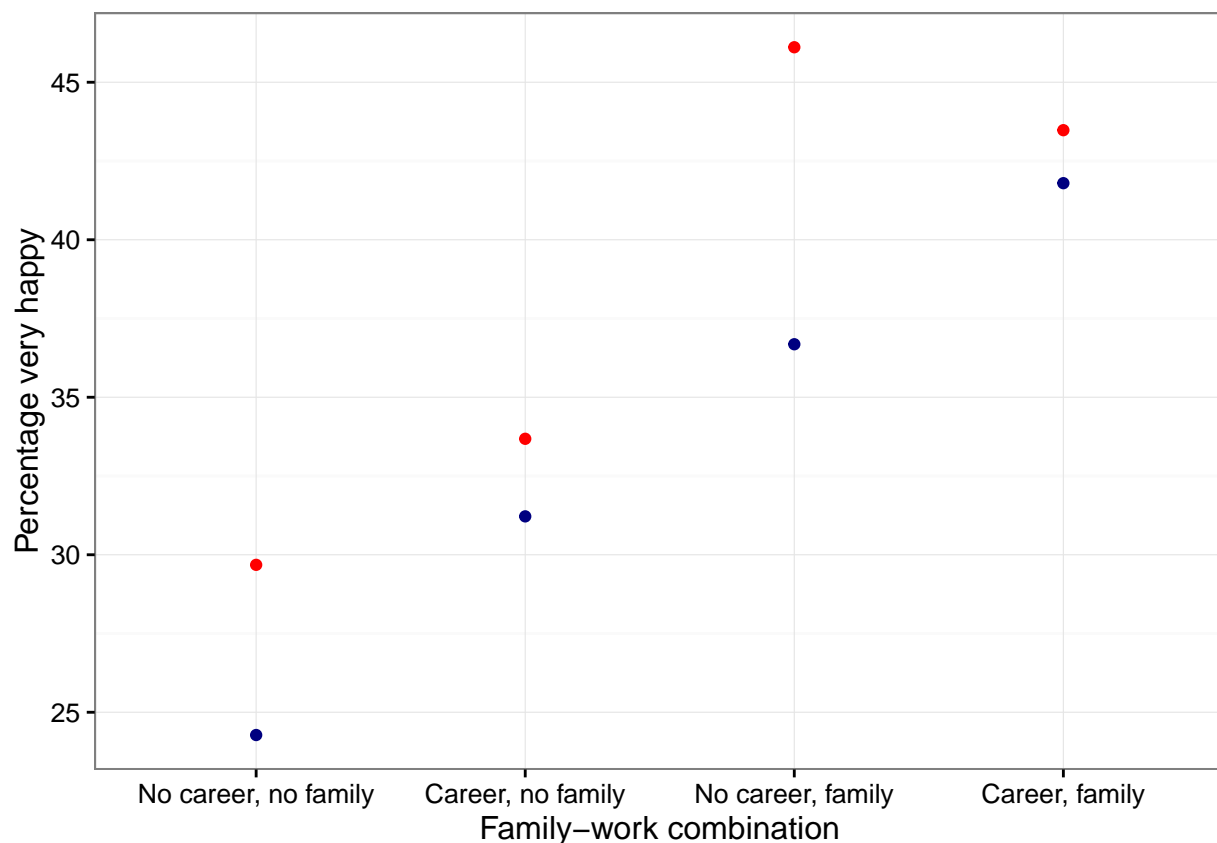

Interaction effects of being married and having a high-income job for working men and women

The correlations shown in figure (x)-(y) are also dependend on omitted factors. To control for some of the confounding factors that are observable, we also replicate a linear probability model by Bertrand (2013)

which estimates the interaction effect of being married and having a high paid job (career) on the binary variable being very happy. While Bertrand (2013) limits her analysis to college-educated women who are working, we also compare these findings to college educated men. The model controls for age, age-squared, the survey year, race and decade of birth.

Figure 5 shows the effect of the interaction term on the probability of being very happy. On the left-hand side the effect of marriage on life satisfaction is stronger for women who do not have a high paying job. For women who have a high paying job, however, the effect of marriage on life satisfaction is much weaker. Although this difference is not significant at the 5% level, the career effect is much stronger than for men, where having a career or not hardly influences the effect of marriage on happiness.

Figure 5: Interaction effects of being married and high income job on life satisfaction

```
z$working_ft <- as.numeric(z$working_ft)

z$working_pt <- as.numeric(z$working_pt)


M1a <- lm(vhappy ~ career*married + age + agesq + as.factor(year) + as.factor(race) + as.factor(bdec),

        data = subset(z, sex==2 & educat == 4))

M2a <- lm(vhappy ~ career*married + age + agesq + as.factor(year) + as.factor(race) + as.factor(bdec),

        data = subset(z, sex==1 & educat == 4))



# compare men and women


require("interplot")

require("gridExtra")


interM1 <- interplot(M1a, var1 = "married", var2 = "career") +

  ggtitle("Working women") +

  xlab("career") +

  ylab("effect of marriage on life satisfaction") +

  expand_limits(y=c(0.05,0.3)) +

  theme_bw()


interM2 <- interplot(M2a, var1 = "married", var2 = "career") +
```

```
  ggtitle("Working men") +

  xlab("career") +

  expand_limits(y=c(0.05,0.3)) +

  theme_bw()


grid.arrange(interM1, interM2, ncol = 2)
```
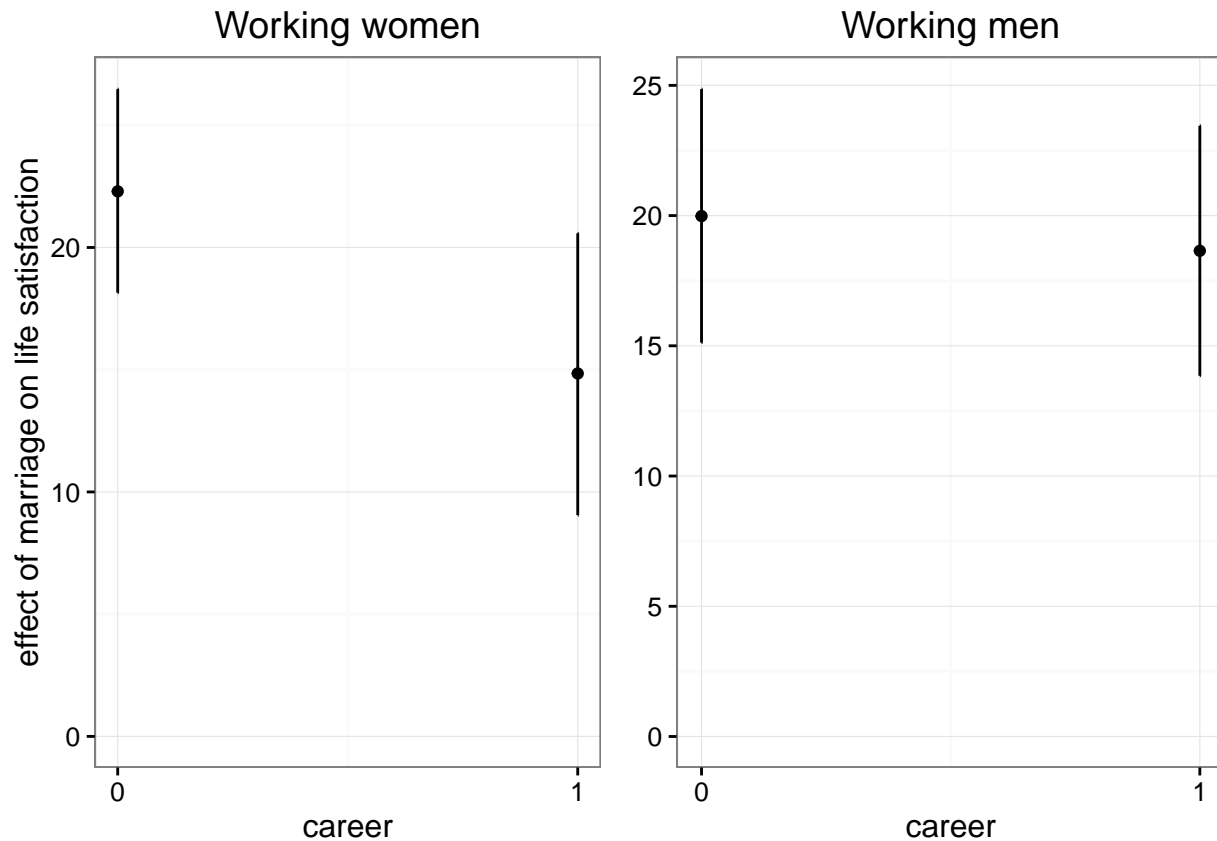

Figure 6: Interaction effects of family-career for working men and women

```
M3 <- lm(vhappy ~ career*family + age + agesq + as.factor(year) + as.factor(race) + as.factor(bdec),
         data = subset(z, sex==2 & educat == 4))
M4 <- lm(vhappy ~ career*family + age + agesq + as.factor(year) + as.factor(race) + as.factor(bdec),
         data = subset(z, sex==1 & educat == 4))


# Compare men and women (Career-Family interaction)


interM3 <- interplot(M3, var1 = "family", var2 = "career") +
```
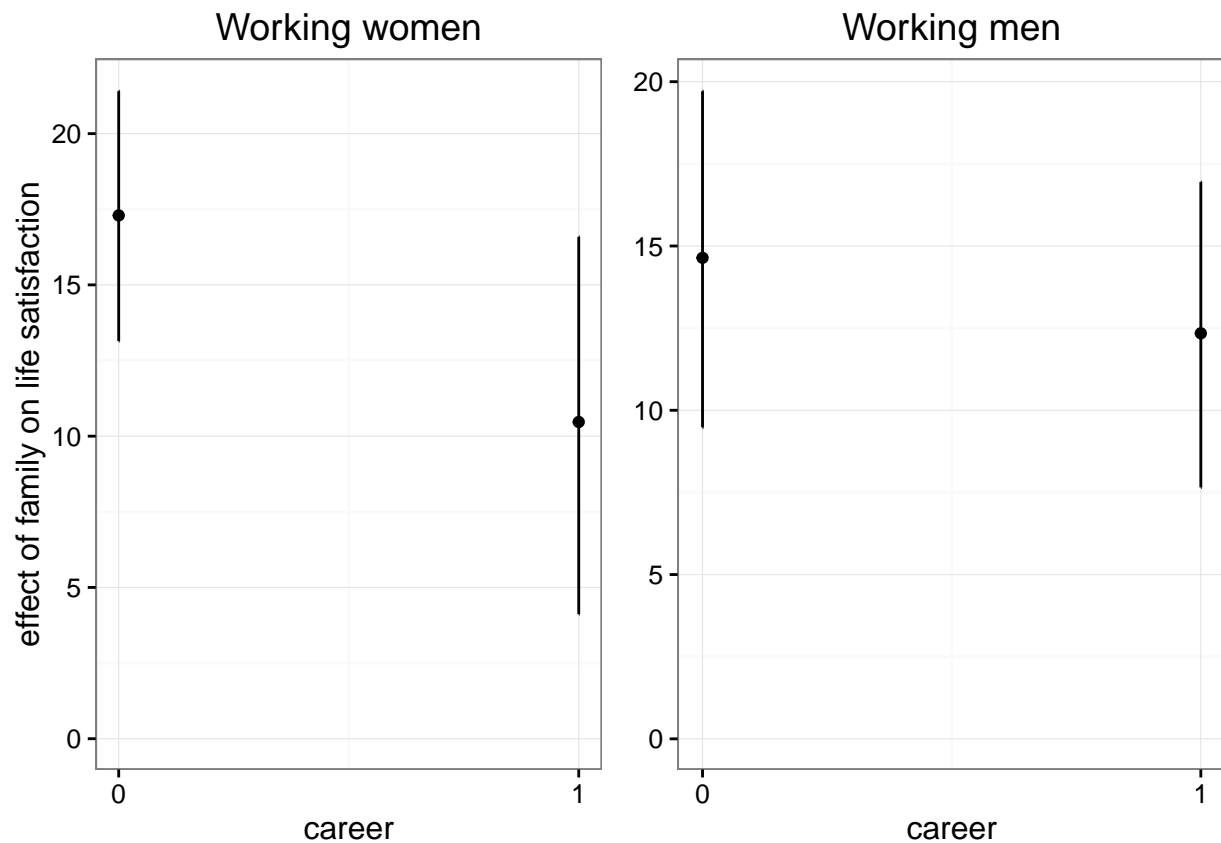
```
  ggtitle("Working women") +

  xlab("career") +

  ylab("effect of family on life satisfaction") +

  expand_limits(y=c(0.05,0.3)) +

  theme_bw()


interM4 <- interplot(M4, var1 = "family", var2 = "career") +

  ggtitle("Working men") +

  xlab("career") +

  expand_limits(y=c(0.05,0.3)) +

  theme_bw()


grid.arrange(interM3, interM4, ncol = 2)
```

Software and packages used for the analysis

The analysis is done in R (R Core Team 2015b) with the use of the following packages: "ggplot2" (Wickham and Chang 2015), "repmis" (Gandrud 2016), "plyr" (Wickham 2015), "dplyr" (Wickham and Francois 2015), "MASS" (Ripley 2015), "Hmisc" (Harrell 2016), "interplot" (Solt and Hu 2016), "gridExtra" (Auguie 2016), "car" (Fox and Weisberg 2016), "foreign" (R Core Team 2015a), "gmodels" (Warnes et al. 2015), "quantmod" (Ryan 2015) and "reshape" (Wickham 2014).

References

Auguie, Baptiste. 2016. *GridExtra: Miscellaneous Functions for "Grid" Graphics.* https://CRAN.R-project.org/package=gridExtra.

Fox, John, and Sanford Weisberg. 2016. *Car: Companion to Applied Regression.* https://CRAN.R-project.org/package=car.

Gandrud, Christopher. 2016. *Repmis: Miscellaneous Tools for Reproducible Research.* https://CRAN.R-project.org/package=repmis.

Harrell, Frank E, Jr. 2016. *Hmisc: Harrell Miscellaneous.* https://CRAN.R-project.org/package=Hmisc.

R Core Team. 2015a. *Foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, DBase, .* https://CRAN.R-project.org/package=foreign.

———. 2015b. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ripley, Brian. 2015. *MASS: Support Functions and Datasets for Venables and Ripley's MASS.* https://CRAN.R-project.org/package=MASS.

Ryan, Jeffrey A. 2015. *Quantmod: Quantitative Financial Modelling Framework.* https://CRAN.R-project.org/package=quantmod.

Solt, Frederick, and Yue Hu. 2016. *Interplot: Plot the Effects of Variables in Interaction Terms.* https://CRAN.R-project.org/package=interplot.

Warnes, Gregory R., Ben Bolker, Thomas Lumley, Randall C Johnson. Contributions from Randall C. Johnson are Copyright SAIC-Frederick, Inc. Funded by the Intramural Research Program, of the NIH,

National Cancer Institute, and Center for Cancer Research under NCI Contract NO1-CO-12400. 2015. *Gmodels: Various R Programming Tools for Model Fitting.* https://CRAN.R-project.org/package=gmodels.

Wickham, Hadley. 2014. *Reshape: Flexibly Reshape Data.* https://CRAN.R-project.org/package=reshape.

———. 2015. *Plyr: Tools for Splitting, Applying and Combining Data.* https://CRAN.R-project.org/package=plyr.

Wickham, Hadley, and Winston Chang. 2015. *Ggplot2: An Implementation of the Grammar of Graphics.* https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, and Romain Francois. 2015. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.