Replication code

Alexey Bessudnov

15 July, 2019

This is a replication code to reproduce the results in Bessudnov, Alexey & Andrey Shcherbak, A. 2018. "Ethnic hierarchy in the Russian labour market: A field experiment". SocArXiv. October 16. doi:10.31235/osf.io/2qzus.

Install and attach packages

```
library(tidyverse)
library(knitr)
library(survey)
library(srvyr)
library(stargazer)
```

Name recognition survey (table 2)

```
Names <- read.csv("SurveyPublic.csv")
Names <- Names %>%
        filter(loc != "abroad") %>%
        filter(ethnLabel != "Bashkir") %>%
        filter(ethnLabel != "Belarusian") %>%
        filter(ethnLabel != "Dagestani") %>%
        filter(ethnLabel != "Estonian") %>%
        filter(ethnLabel != "Kazakh") %>%
        filter(ethnLabel != "Kyrgyz") %>%
        filter(ethnLabel != "Moldovan")
Names %>%
        group_by(ethnLabel) %>%
        summarise(
                Correct = mean(correct, na.rm = TRUE) * 100,
                CorrectBroad = mean(correct.broad, na.rm = TRUE) * 100,
                NotRussian = mean(notRussian, na.rm = TRUE) * 100
        ) %>%
        arrange(desc(Correct)) %>%
        kable(col.names = c("Ethnic group", "% correct", "% broadly correct", "% not Russian"),
        digits = 0)
```

Ethnic group	% correct	% broadly correct	% not Russian
Georgian	91	98	100
Armenian	90	96	100
Russian	88	90	12
Ukrainian	82	92	95
Jewish	72	84	99
Tatar	57	90	99

Ethnic group	% correct	% broadly correct	% not Russian
German	42	62	85
Latvian	35	65	100
Lithuanian	22	73	100
Chechen	20	83	99
Uzbek	19	91	100
Azeri	16	90	100
Tajik	12	84	99

Open the data from the experiment and recode variables

```
Vacancies <- read_csv("VacanciesPublic.csv")

# Converting variables into factors
Vacancies$RA <- factor(Vacancies$RA)
Vacancies$city <- factor(Vacancies$city, levels = c("Moscow", "St Petersburg", "Kazan", "Ufa"))
Vacancies$city2 <- factor(Vacancies$city2, levels = c("Moscow/St Petersburg", "Kazan/Ufa"))
Vacancies$occupation <- factor(Vacancies$occupation, levels = c("cook", "salesperson",
                                                                "sales manager", "programmer"))
Vacancies$gender <- factor(Vacancies$gender, levels = c("female", "male"))
Vacancies$ethnicity <- factor(Vacancies$ethnicity,
        levels = c("Russian", "Jewish", "Ukrainian",
        "German", "Lithuanian", "Latvian", "Georgian", "Armenian", "Tatar",
        "Bashkir", "Chechen", "Azeri", "Tajik", "Uzbek"))
Vacancies$ethnCollapsed <- factor(Vacancies$ethnCollapsed,
        levels = c("Russian", "Jewish", "Ukrainian", "German", "Latvian/Lithuanian",
                   "Tatar", "Tajik/Uzbek", "Chechen/Azeri",  "Armenian", "Georgian"))
Vacancies$ethnGroup2 <- factor(Vacancies$ethnGroup2,
        levels = c("European", "Southern"))
Vacancies$website <- factor(Vacancies$website)
Vacancies$candidateID <- factor(Vacancies$candidateID)
```

Total number of job applications and contact rates (section 5.1)

```
Vacancies %>%
        filter(ethnicity != "Bashkir") %>%
       nrow() %>%
        kable()
```

x
9607

```
Vacancies %>%
        filter(ethnicity != "Bashkir") %>%
        summarise(
                overall = sum(response) / n(),
                onPhone = sum(phoneContact) / n(),
```

```
                onWebsite = sum(websiteContact) / n()
        ) %>%
        kable(digits = 2)
```

overall	onPhone	onWebsite
0.36	0.21	0.23

Contact rates by ethnic group and location (table 4)

```
descTable <- Vacancies %>%
        filter(ethnicity != "Bashkir") %>%
#         Using the srvyr package that lets us work with the survey functions
# within the tidyverse structures
        as_survey_design(ids = candidateID) %>%
        group_by(city2, ethnCollapsed) %>%
        summarise(
                # Total number of applications (with standard error that is unnecessary)
                n_app = survey_total(),
                # Total number of responses (with standard error that is unnecessary)
                n_response = survey_total(response),
                propContact = survey_mean(response, vartype = "ci")
        ) %>%
        as_tibble()

# odds of receiving response for Russians in Moscow/St Peterbsurg
oddsRusMSP <- descTable %>%
        filter(ethnCollapsed == "Russian" & city2 == "Moscow/St Petersburg") %>%
        summarise(
           odds =   n_response / (n_app - n_response)
        ) %>%
        as.numeric()
# odds of receiving response for Russians in Kazan/Ufa
oddsRusKU <- descTable %>%
        filter(ethnCollapsed == "Russian" & city2 == "Kazan/Ufa") %>%
        summarise(
           odds =   n_response / (n_app - n_response)
        ) %>%
        as.numeric()

# proportion response for Russians in Moscow/StPeterbsurg
propRusMSP <- descTable %>%
        filter(ethnCollapsed == "Russian" & city2 == "Moscow/St Petersburg") %>%
        select(propContact) %>%
        as.numeric()

# proportion response for Russians in Kazan/Ufa
propRusKU <- descTable %>%
        filter(ethnCollapsed == "Russian" & city2 == "Kazan/Ufa") %>%
        select(propContact) %>%
        as.numeric()
```

```
# Updating the table
descTable <- descTable %>%
        select(-c(n_response_se, n_app_se))
```

Moscow and St Petersburg

```
# Moscow and St Petersburg
descTable %>%
        filter(city2 == "Moscow/St Petersburg") %>%
        mutate(oddsRus = oddsRusMSP) %>%
        mutate(propRus = propRusMSP) %>%
        mutate(cbkRatio = propRus / propContact) %>%
        mutate(or = (n_response / (n_app - n_response)) / oddsRus) %>%
        select(-c(city2, propRus, oddsRus)) %>%
        arrange(desc(or)) %>%
        kable(digits = 2)
```

ethnCollapsed	n_app	n_response	propContact	propContact_low	propContact_upp	cbkRatio	or
Russian	616	254	0.41	0.35	0.47	1.00	1.00
Ukrainian	559	220	0.39	0.34	0.45	1.05	0.92
Jewish	604	237	0.39	0.35	0.44	1.05	0.92
German	642	232	0.36	0.32	0.40	1.14	0.81
Latvian/Lithuanian	551	185	0.34	0.29	0.38	1.23	0.72
Tajik/Uzbek	570	159	0.28	0.22	0.34	1.48	0.55
Tatar	610	170	0.28	0.23	0.32	1.48	0.55
Chechen/Azeri	598	165	0.28	0.23	0.32	1.49	0.54
Armenian	610	163	0.27	0.22	0.31	1.54	0.52
Georgian	549	142	0.26	0.21	0.30	1.59	0.50

Kazan and Ufa

```
# Kazan and Ufa
descTable%>%
        filter(city2 == "Kazan/Ufa") %>%
        mutate(oddsRus = oddsRusKU) %>%
        mutate(propRus = propRusKU) %>%
        mutate(cbkRatio = propRus / propContact) %>%
        mutate(or = (n_response / (n_app - n_response)) / oddsRus) %>%
        select(-c(city2, propRus, oddsRus)) %>%
        arrange(desc(or)) %>%
        kable(digits = 2)
```

ethnCollapsed	n_app	n_response	propContact	propContact_low	propContact_upp	cbkRatio	or
Jewish	384	187	0.49	0.42	0.55	0.89	1.24
German	369	167	0.45	0.37	0.53	0.96	1.08
Russian	402	174	0.43	0.38	0.49	1.00	1.00
Tatar	343	147	0.43	0.36	0.49	1.01	0.98
Ukrainian	365	155	0.42	0.38	0.47	1.02	0.97
Tajik/Uzbek	373	150	0.40	0.35	0.46	1.08	0.88

ethnCollapsed	n_app	n_response	propContact	propContact_low	propContact_upp	cbkRatio	or
Armenian	378	148	0.39	0.34	0.44	1.11	0.84
Georgian	368	144	0.39	0.31	0.47	1.11	0.84
Latvian/Lithuanian	355	138	0.39	0.34	0.44	1.11	0.83
Chechen/Azeri	361	138	0.38	0.32	0.44	1.13	0.81

n_app: the number of job applications sent;

n_response: the number of positive responses received;

propContact: proportion contacted (n_response / n_app);

propContact_low: lower bound of the 95% CI for prop_Contact;

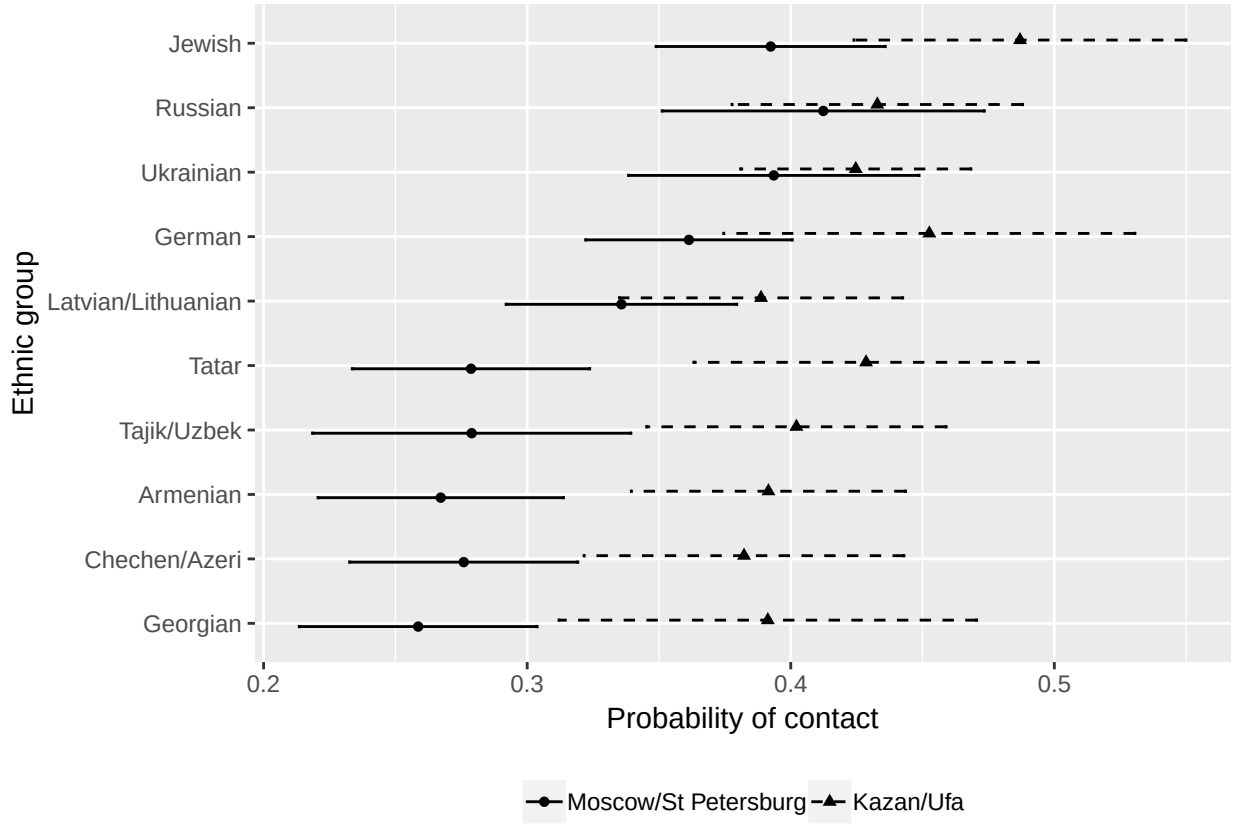
propContact_upp: upper bound of the 95% CI for prop_Contact;

cbkRatio: calllback ratio (propRus / propContact);

or: odds ratio ((n_response / n_nonresponse) / (n_responseRus / n_nonresponseRus)).

Contact rates by ethnic group and location: chart (figure 1)

```
Vacancies %>%
        filter(ethnicity != "Bashkir") %>%
        as_survey_design(ids = candidateID) %>%
        group_by(city2, ethnCollapsed) %>%
        summarise(
                propContact = survey_mean(response, vartype = "ci")
        ) %>%
        arrange(desc(propContact)) %>%
        ggplot(aes(x = reorder(ethnCollapsed, propContact), y = propContact,
                        ymin = propContact_low, ymax = propContact_upp, linetype = city2)) +
        geom_point(position = position_dodge(width = 0.2), aes(shape = city2)) +
        geom_errorbar(position = position_dodge(width = 0.2), width = 0.1) +
        coord_flip() +
        ylab("Probability of contact") +
        xlab("Ethnic group") +
        labs(shape="City") +
        scale_linetype_manual(name = "City", values = 1:2) +
        scale_shape_manual(name = "City", values = 16:17) +
        theme(legend.position = "bottom") +
        theme(legend.title=element_blank())
```



Linear probability models for contact (table 5)

```
# Moscow and St Petersburg
Vacancies.MSP <- Vacancies %>%
        filter(city2 == "Moscow/St Petersburg") %>%
        as_survey_design(ids = candidateID)
Vacancies.KU <- Vacancies %>%
        filter(city2 == "Kazan/Ufa") %>%
        as_survey_design(ids = candidateID)

# Estimating the models

m1 <- svyglm(response ~ ethnCollapsed + gender + occupation + city + website + RA,
             design = Vacancies.MSP)
m2 <- svyglm(response ~ ethnCollapsed + gender + occupation + city + website + RA,
             design = Vacancies.KU)

stargazer(m1, m2, omit.stat = c("rsq", "adj.rsq", "f", "ser", "bic", "aic", "ll"),
          column.labels = c("Moscow/St Petersburg", "Kazan/Ufa"),
          omit = c("gender", "occupation", "city", "website", "RA", "Constant"),
          covariate.labels = c("Jewish", "Ukrainian", "German", "Latvian/Lithuanian",
                               "Tatar", "Tajik/Uzbek", "Azerbaijani/Chechen","Armenian",
                               "Georgian"),
          digits = 2,
```

```
          star.cutoffs = c(0.05, 0.01, 0.001))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Jul 15, 2019 - 16:33:56

Table 6:

	Dependent variable:	
	response	
	Moscow/St Petersburg	Kazan/Ufa
	(1)	(2)
Jewish	−0.02	0.06
	(0.04)	(0.04)
Ukrainian	−0.02	0.002
	(0.03)	(0.04)
German	−0.05	0.04
	(0.03)	(0.04)
Latvian/Lithuanian	−0.07*	−0.04
	(0.03)	(0.04)
Tatar	−0.13***	0.005
	(0.03)	(0.04)
Tajik/Uzbek	−0.13***	−0.03
	(0.03)	(0.04)
Azerbaijani/Chechen	−0.13***	−0.04
	(0.03)	(0.04)
Armenian	−0.14***	−0.03
	(0.03)	(0.04)
Georgian	−0.15***	−0.03
	(0.03)	(0.04)
Observations	5,909	3,698
Note:	*p<0.05; **p<0.01; ***p<0.001	

Linear probability models. All the models control for gender, occupation, city, website, and research assistant's name. Cluster-robust standard errors applied (clustered by applicant's name). Ethnic Russians are the reference group.

Interaction between ethnicity and gender (table 6)

```
# Gender
```

```
mGender1 <- svyglm(response ~ ethnGroup2  * gender + occupation + city + website + RA,
                   design = Vacancies.MSP)
mGender2 <- svyglm(response ~ ethnGroup2  * gender + occupation + city + website + RA,
                   design = Vacancies.KU)

stargazer(mGender1, mGender2, omit.stat = c("rsq", "adj.rsq", "f", "ser", "bic", "aic", "ll"),
          column.labels = c("Moscow/St Petersburg", "Kazan/Ufa"),
          omit = c("occupation", "city", "website", "RA", "Constant"),
          covariate.labels = c("Southern", "male", "Southern:male"),
          digits = 2,
          star.cutoffs = c(0.05, 0.01, 0.001))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Jul 15, 2019 - 16:33:56

Table 7:

	Dependent variable:	
	response	
	Moscow/St Petersburg	Kazan/Ufa
	(1)	(2)
Southern	−0.07***	−0.02
	(0.02)	(0.02)
male	0.0001	0.01
	(0.02)	(0.03)
Southern:male	−0.07*	−0.03
	(0.03)	(0.03)
Observations	5,909	3,698
Note:	*p<0.05; **p<0.01; ***p<0.001	

Probability of getting an explicit rejection (table 7)

```
# Subset those with response = 0 only
Rejected <- Vacancies %>%
  filter(response == 0)

# Setting survey objects
# Moscow and St Petersburg
Rejected.MSP <- Rejected %>%
        filter(city2 == "Moscow/St Petersburg") %>%
        as_survey_design(ids = candidateID)
Rejected.KU <- Rejected %>%
        filter(city2 == "Kazan/Ufa") %>%
        as_survey_design(ids = candidateID)

# Contact on the phone
```

```
mRejected1 <- svyglm(rejected ~ ethnCollapsed + website + gender + occupation  + city + RA,
                     design = Rejected.MSP)
mRejected2 <- svyglm(rejected ~ ethnCollapsed + website + gender + occupation  + city + RA,
                     design = Rejected.KU)


stargazer(mRejected1, mRejected2, omit.stat = c("rsq", "adj.rsq", "f", "ser", "bic", "aic", "ll"),
          column.labels = c("Moscow/St Petersburg", "Kazan/Ufa"),
          omit = c("website", "gender", "occupation", "city", "RA", "Constant"),
          covariate.labels = c("Jewish", "Ukrainian", "German", "Latvian/Lithuanian",
                               "Tatar", "Tajik/Uzbek", "Azerbaijani/Chechen",
                               "Armenian", "Georgian"),
          digits = 2,
          star.cutoffs = c(0.05, 0.01, 0.001))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Jul 15, 2019 - 16:33:56

Contact on the phone and on the websites (table 8)

```
# Subset to those with response = 1 only
Response <- Vacancies %>%
  filter(response == 1)

# Setting survey objects
# Moscow and St Petersburg
Response.MSP <- Response %>%
        filter(city2 == "Moscow/St Petersburg") %>%
        as_survey_design(ids = candidateID)
Response.KU <- Response %>%
        filter(city2 == "Kazan/Ufa") %>%
        as_survey_design(ids = candidateID)

# Contact on the phone
mPhone1 <- svyglm(phoneContact ~ ethnCollapsed + website + gender + occupation  + city + RA,
                  design = Response.MSP)
mPhone2 <- svyglm(phoneContact ~ ethnCollapsed + website + gender + occupation  + city + RA,
                  design = Response.KU)


stargazer(mPhone1, mPhone2, omit.stat = c("rsq", "adj.rsq", "f", "ser", "bic", "aic", "ll"),
          column.labels = c("Moscow/St Petersburg", "Kazan/Ufa"),
          omit = c("website", "gender", "occupation", "city", "RA", "Constant"),
          covariate.labels = c("Jewish", "Ukrainian", "German", "Latvian/Lithuanian",
                               "Tatar", "Tajik/Uzbek", "Azerbaijani/Chechen",
                               "Armenian", "Georgian"),
          digits = 2,
          star.cutoffs = c(0.05, 0.01, 0.001))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Jul 15, 2019 - 16:33:57

Table 8:

	Dependent variable:	
	rejected	
	Moscow/St Petersburg	Kazan/Ufa
	(1)	(2)
Jewish	−0.02	0.04
	(0.02)	(0.03)
Ukrainian	−0.005	0.05
	(0.03)	(0.04)
German	−0.03	0.02
	(0.03)	(0.03)
Latvian/Lithuanian	0.02	0.03
	(0.03)	(0.03)
Tatar	−0.004	0.03
	(0.02)	(0.04)
Tajik/Uzbek	0.06**	−0.004
	(0.02)	(0.03)
Azerbaijani/Chechen	0.08**	0.02
	(0.03)	(0.03)
Armenian	0.03	0.07
	(0.02)	(0.04)
Georgian	0.05*	0.05
	(0.02)	(0.04)
Observations	3,982	2,150
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 9:

	Dependent variable:	
	phoneContact	
	Moscow/St Petersburg	Kazan/Ufa
	(1)	(2)
Jewish	-0.11^{*}	-0.07
	(0.05)	(0.04)
Ukrainian	-0.07	0.01
	(0.05)	(0.05)
German	-0.12^{*}	0.06
	(0.06)	(0.04)
Latvian/Lithuanian	-0.17^{**}	-0.03
	(0.06)	(0.05)
Tatar	-0.18^{**}	-0.02
	(0.06)	(0.07)
Tajik/Uzbek	-0.22^{***}	-0.10
	(0.05)	(0.05)
Azerbaijani/Chechen	-0.19^{***}	-0.01
	(0.05)	(0.06)
Armenian	-0.16^{**}	-0.02
	(0.06)	(0.04)
Georgian	-0.17^{**}	-0.02
	(0.06)	(0.05)
Observations	1,927	1,548
Note:	*p<0.05; **p<0.01; ***p<0.001	

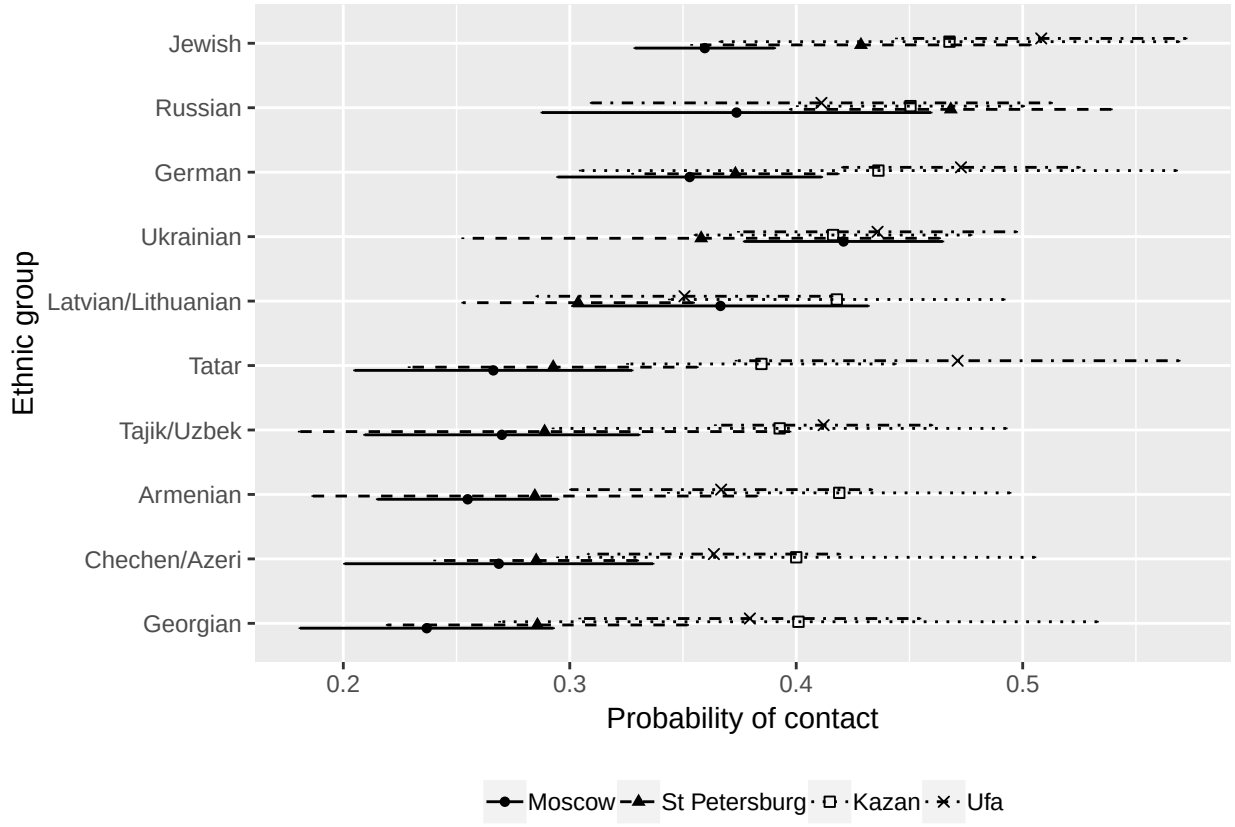
Ethnicity of the employers in Kazan and Ufa (Discussion)

```
Vacancies %>%
        filter(!is.na(notRussianEmp)) %>%
        count(city, notRussianEmp) %>%
        group_by(city) %>%
        mutate(freq = n / sum(n)) %>%
        kable(digits = 2)
```

city	notRussianEmp	n	freq
Moscow	0	861	0.93
Moscow	1	65	0.07
St Petersburg	0	744	0.96
St Petersburg	1	33	0.04
Kazan	0	480	0.72
Kazan	1	188	0.28
Ufa	0	489	0.78
Ufa	1	140	0.22

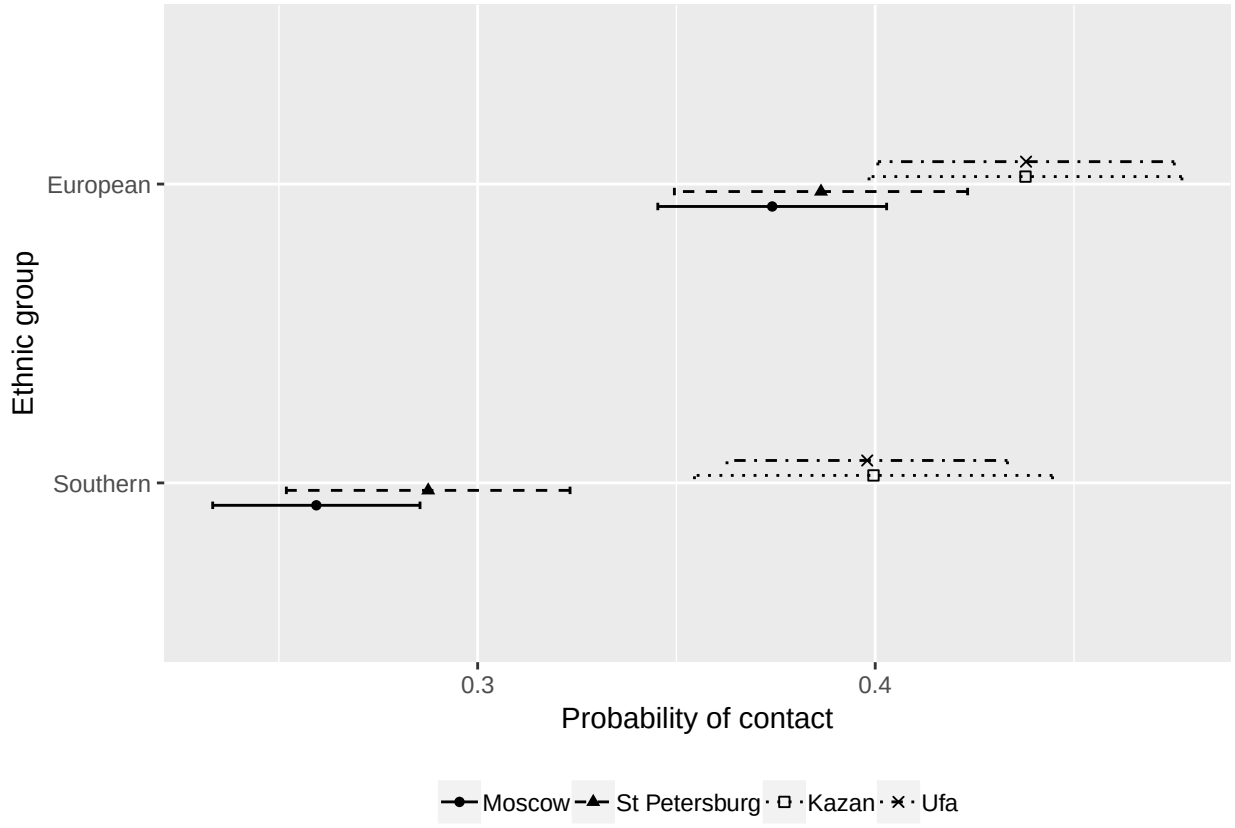
Contact rates by ethnic group in four cities (Appendix)

```
Vacancies %>%
        filter(ethnicity != "Bashkir") %>%
        as_survey_design(ids = candidateID) %>%
        group_by(city, ethnCollapsed) %>%
        summarise(
                propContact = survey_mean(response, vartype = "ci")
        ) %>%
        arrange(desc(propContact)) %>%
        ggplot(aes(x = reorder(ethnCollapsed, propContact), y = propContact,
                   ymin = propContact_low, ymax = propContact_upp, linetype = city)) +
        geom_point(position = position_dodge(width = 0.2), aes(shape = city)) +
        geom_errorbar(position = position_dodge(width = 0.2), width = 0.1) +
        coord_flip() +
        ylab("Probability of contact") +
        xlab("Ethnic group") +
        labs(shape="City") +
        scale_linetype_manual(name = "City",
                              values = 1:4) +
        scale_shape_manual(name = "City",
                           values = c(16:17, 0, 4)) +
        theme(legend.position = "bottom") +
        theme(legend.title=element_blank())
```



Contact rates by ethnic group (combined) in four cities (Appendix)

```
Vacancies %>%
        filter(ethnicity != "Bashkir") %>%
        as_survey_design(ids = candidateID) %>%
        group_by(city, ethnGroup2) %>%
        summarise(
                propContact = survey_mean(response, vartype = "ci")
        ) %>%
        arrange(desc(propContact)) %>%
        ggplot(aes(x = reorder(ethnGroup2, propContact), y = propContact,
                   ymin = propContact_low, ymax = propContact_upp, linetype = city)) +
        geom_point(position = position_dodge(width = 0.2), aes(shape = city)) +
        geom_errorbar(position = position_dodge(width = 0.2), width = 0.1) +
        coord_flip() +
        ylab("Probability of contact") +
        xlab("Ethnic group") +
        labs(shape="City") +
        scale_linetype_manual(name = "City",
                              values = 1:4) +
        scale_shape_manual(name = "City",
                           values = c(16:17,0,4)) +
        theme(legend.position = "bottom") +
        theme(legend.title=element_blank())
```



Interaction between ethnicity and occupation (Appendix)

```
mOcc1 <- svyglm(response ~ ethnGroup2  * occupation + gender + city + website + RA,
                design = Vacancies.MSP)
mOcc2 <- svyglm(response ~ ethnGroup2  * occupation + gender + city + website + RA,
                design = Vacancies.KU)

stargazer(mOcc1, mOcc2, omit.stat = c("rsq", "adj.rsq", "f", "ser", "bic", "aic", "ll"),
          column.labels = c("Moscow/St Petersburg", "Kazan/Ufa"),
          omit = c("gender", "city", "website", "RA", "Constant"),
          covariate.labels = c("Southern", "salesperson", "sales manager",
                    "programmer", "Southern * salesperson", "Southern * sales manager",
                    "Southern * programmer"),
          digits = 2,
          star.cutoffs = c(0.05, 0.01, 0.001))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Jul 15, 2019 - 16:33:58

Table 11:

	Dependent variable:	
	response	
	Moscow/St Petersburg	Kazan/Ufa
	(1)	(2)
Southern	-0.10^{***}	0.04
	(0.03)	(0.04)
salesperson	-0.05	-0.10^{*}
	(0.03)	(0.04)
sales manager	-0.11^{***}	-0.03
	(0.03)	(0.04)
programmer	-0.09^{**}	0.07
	(0.03)	(0.05)
Southern * salesperson	-0.01	-0.07
	(0.04)	(0.05)
Southern * sales manager	-0.004	-0.11^{*}
	(0.04)	(0.05)
Southern * programmer	0.01	-0.09
	(0.04)	(0.07)
Observations	5,909	3,698
Note:	*p<0.05; **p<0.01; ***p<0.001	