Course: Data Analysis (task № 2)

Student’s Name and Surname: Pavel Drankov

1. Open **salary.dta** file.   
 sysuse salary.dta

* 1. Recode **position** variable from string into numeric.

encode position, generate(num\_position)

1.2. Calculate a new variable which will show the sum of income taxes **in dollars** which is paid monthly to every person (use today’s exchange rate). Take into account that foreigners pay 30% of income taxes and non-foreigners – 13%.

generate dollor\_rub\_ratio = 64

generate dollor\_rub\_ratio = 64

generate tax = .

replace tax = (salary / dollor\_rub\_ratio) \* 0.13 if foreigner == 0

replace tax = (salary / dollor\_rub\_ratio) \* 0.30 if foreigner == 1

tab tax

1.3. Create a new variable which will represent the rank of a person which is calculated on the basis of his/her salary. Person with the highest salary will have rank 1.

egen rnk = rank(salary), field

1.4. Sort the data of the basis of the salary variable so that the person with the highest salary appears in the first row.

gsort -salary

1.5. Save all the changes in **salary.dta** file.

save salary\_modif.dta

2. Open **data\_games.dta** file.

2.1. Deselect form the further analysis the cases which have the value of “**payment**” variable less than 100 or greater than 14000.

drop if payment < 100 | payment > 14000

2.2. Divide the remained cases into 4 groups using the following interval width:

1st group: from 100 to 500;

2nd group: from 501 to 1000;

3rd group: from 1001 to 2000;

4th group: greater than 2000.

Indicate the mode group number.

recode payment (100/500 = 1 "1st group") (501/1000 = 2 "2nd group") (1001/2000 = 3 "3rd group") (2000/max = 4 "4rd group" ), gen(payment\_group)

tab payment\_group

3. Open **revenue.dta** file.

Modify the dataset so as it contains three variables:

- year;

- company;

- new variable which represents the mean revenue of each company for each year.

collapse (sum) year\_revenue=rev, by(year company)

4. Contingency table. Open **data\_games.dta** file.

Create a contingency table between two variables **payment\_method** and **country**. Answer the following questions.   
tab2 payment\_method country

4.1. Which payment method is most frequently used in Russia (RU)?   
 Answer: general

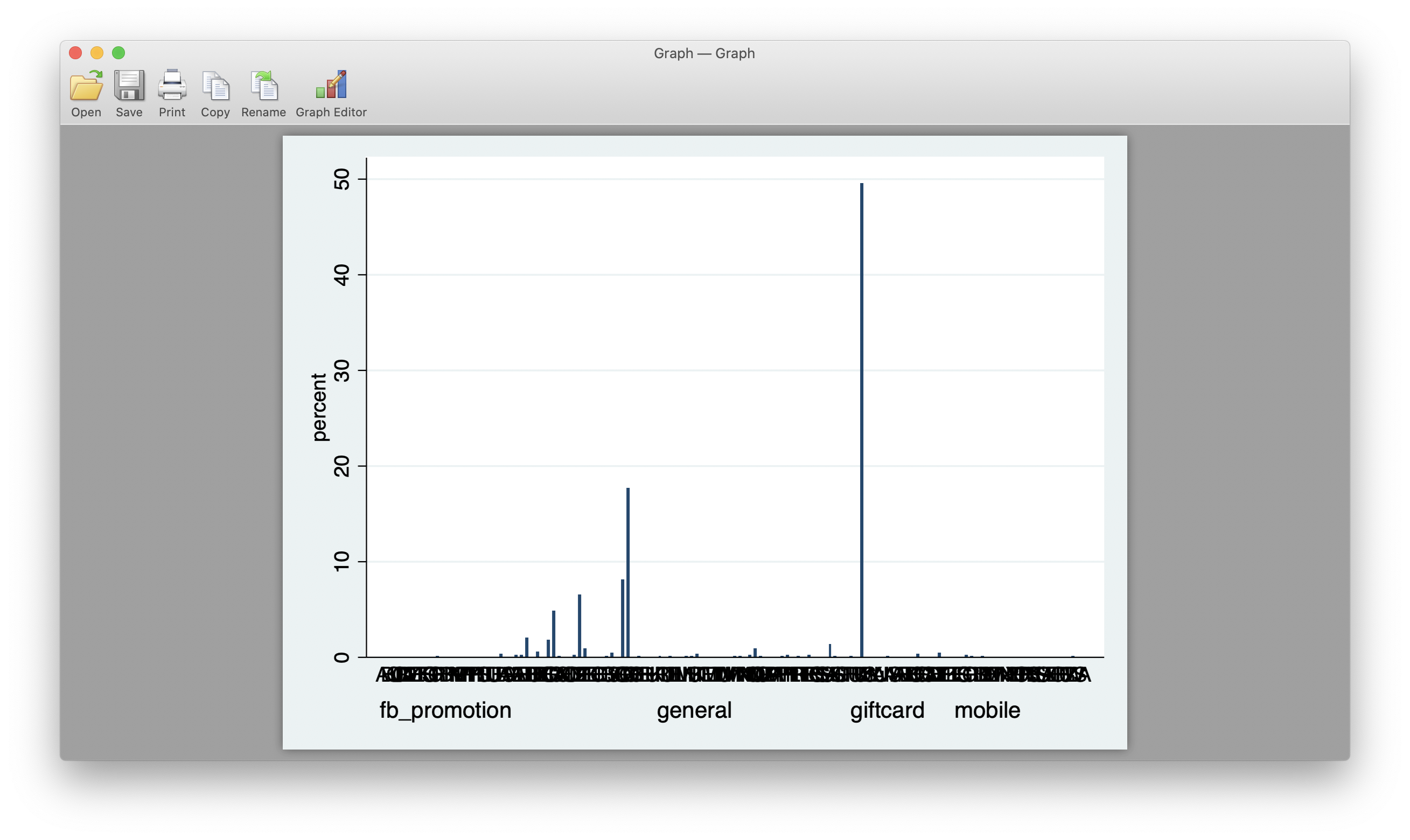
4.2. What is the percentage of payments which were done using a gift card from all the worldwide payments?

Answer: 0.05

4.3. Players from which country have made the highest number of payments using a mobile payment type? Answer: DE, 495  
 drop if payment\_method != 4

tab2 payment\_method country, colsort

4.4. Create a suitable graph to demonstrate the relationship between these two variables.



Please send to amelikyan@hse.ru from your personal e-mail:

- this MS Word file with answers,

- the do-file with the corresponding commands,

- the modified salary.dta file.