Ishmametyev Nikolay

Data analysis in Python

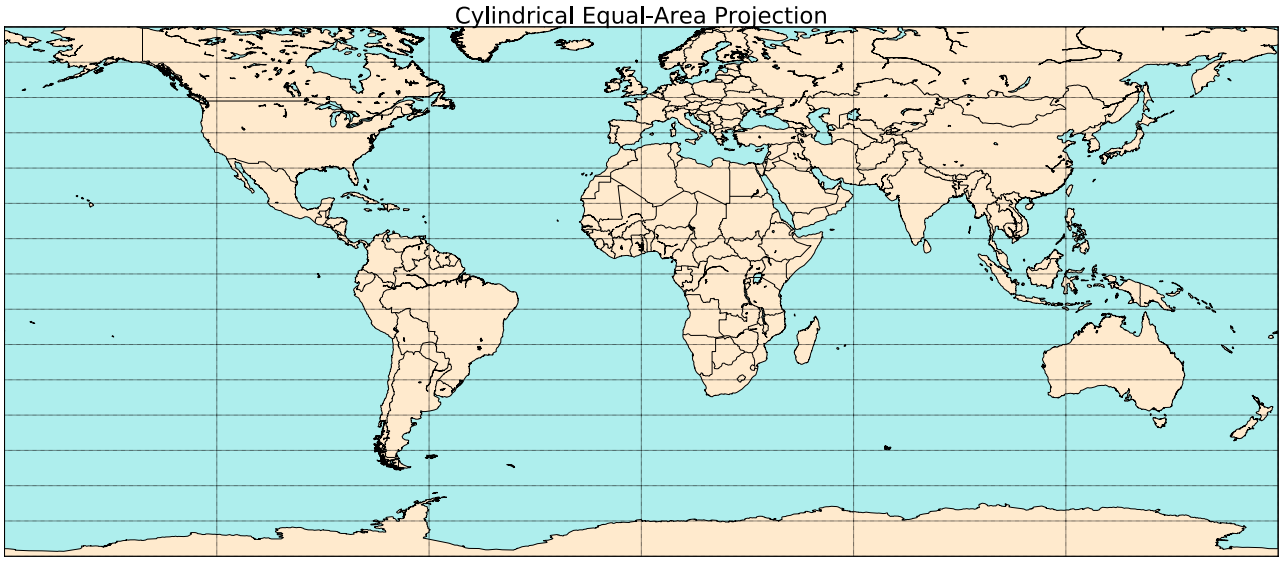
Home assignment 5

Task 1

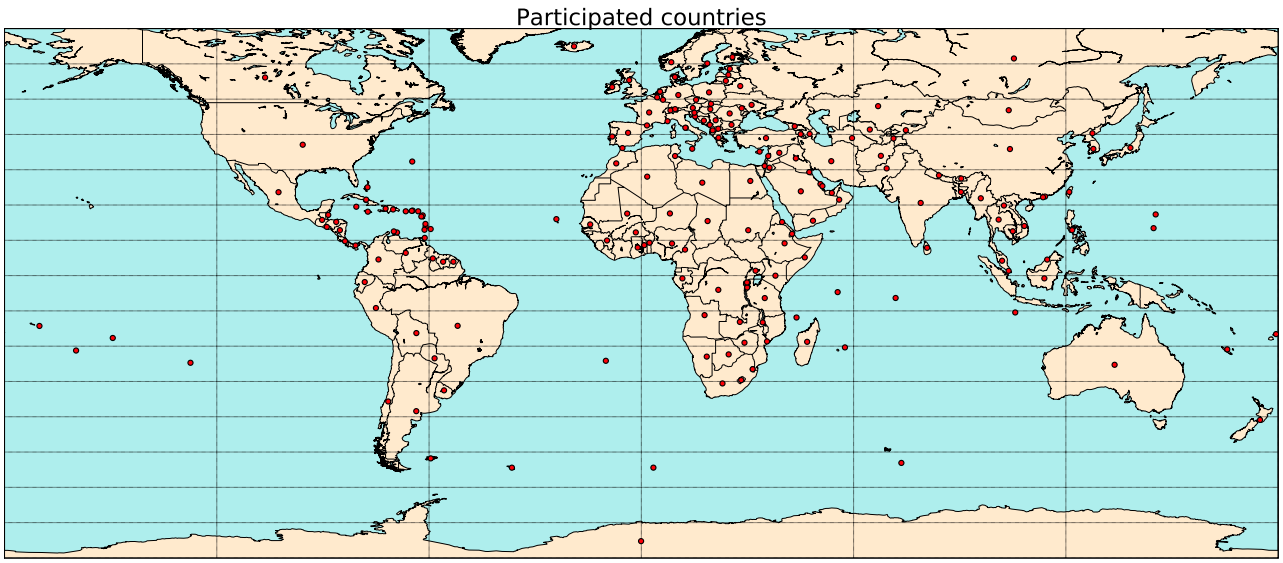
1. Zip file with .csv documents was downloaded via *request* library. Then two files .csv was extracted from zip and were written down.

Answer on the first Task is:
Number of questions 154
Number of respondents 51392

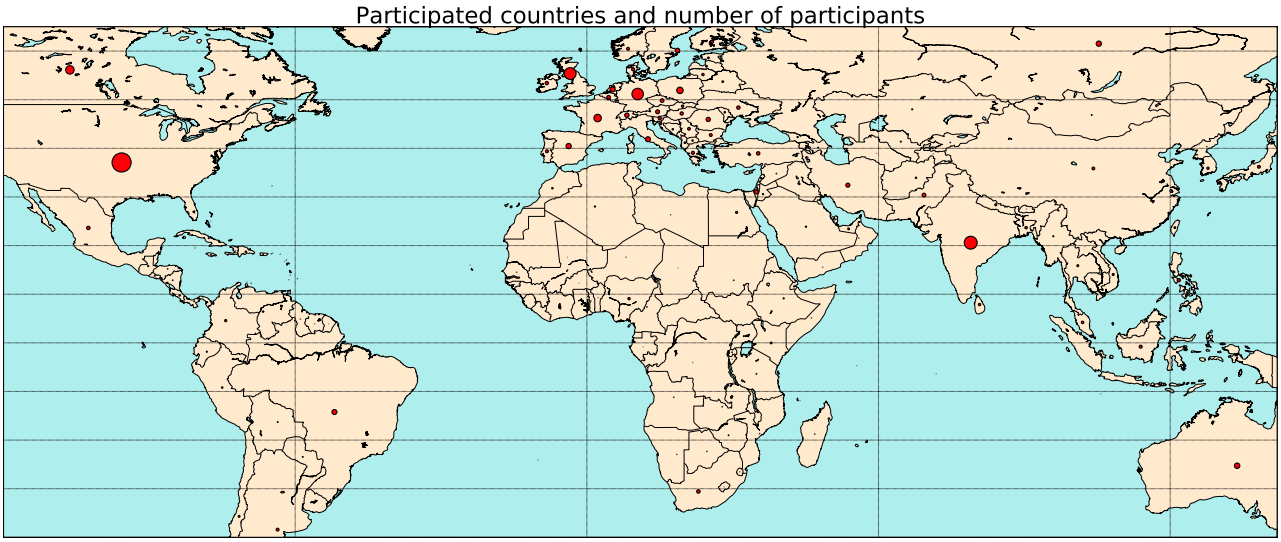
2. The world map was drawn via library BaseMap in Python. Picture is shown below.



3. Geo-positions of countries which participated in survey were extracted from csv on the web-site *https://developers.google.com/public-data/docs/canonical/countries_csv*. In the centers of these countries are placed red dots. Picture is presented below.



4. In this subtask we need to draw dots accordingly to number of participants from particular country. Result is shown below.



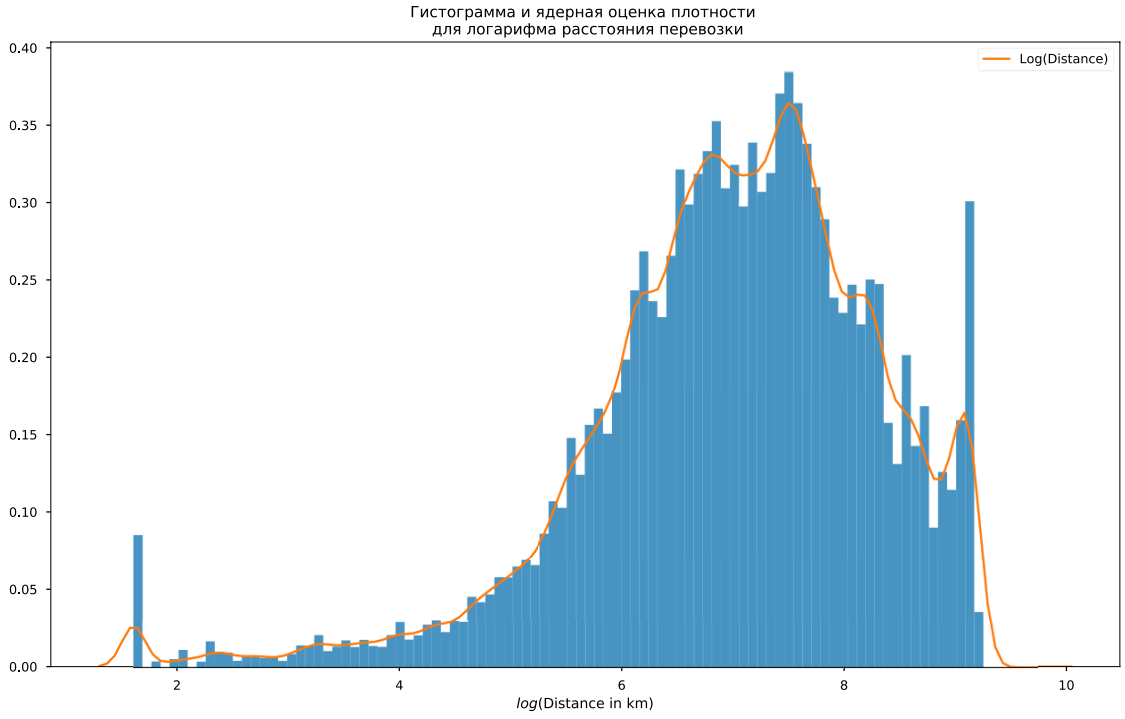
Task 2

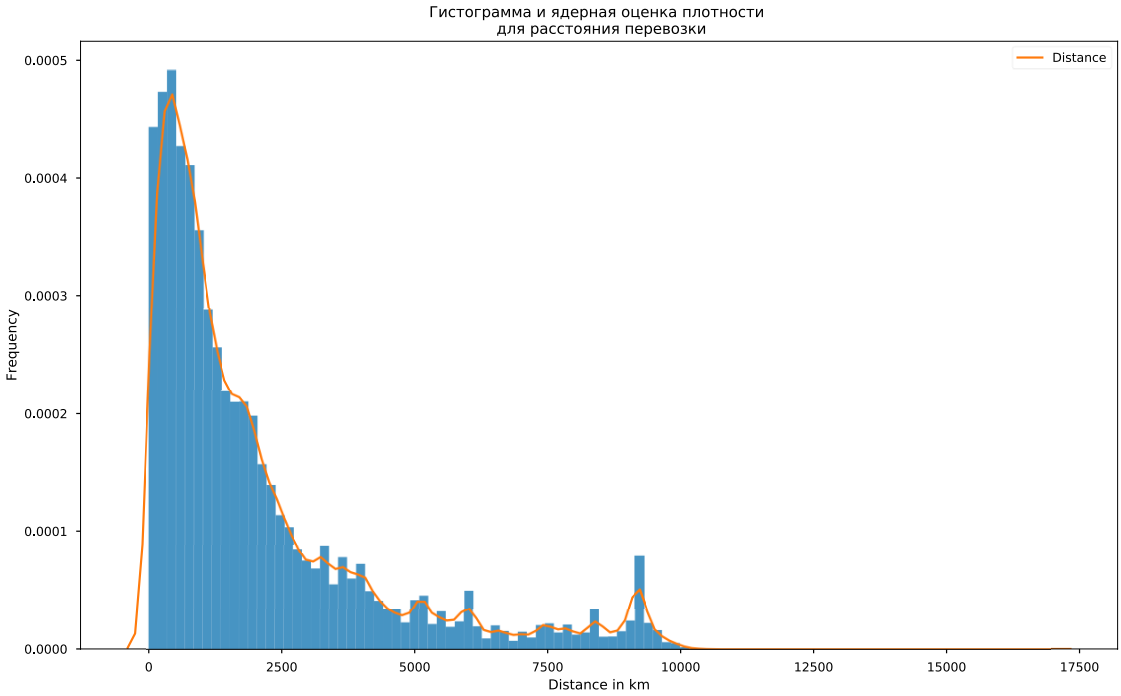
1. I've created two DataFrames from csv files. Heads of then are shown below.

	date_priem	fr_code	sto_code	stn_code	dist
0	31.08.2012	1100	61400	3010	1944
1	01.08.2012	1100	61360	59250	431
2	29.08.2012	1100	81530	60530	1443

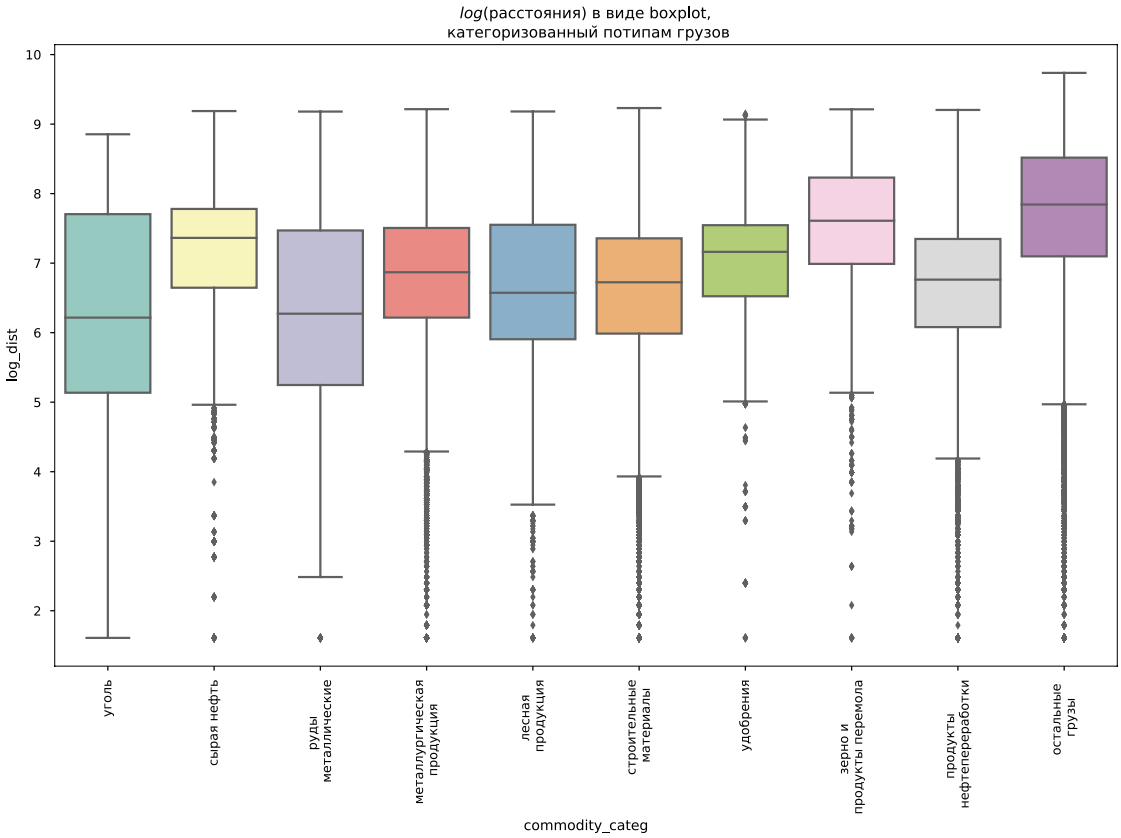
	stshortname	stname	stcode
0	ВЫБОРГ-ПЕРЕВ	ВЫБОРГ-ПЕРЕВАЛКА	2340
1	КАЛАШНИКОВО	КАЛАШНИКОВО	6230
2	ДОБЫВАЛОВО	ДОБЫВАЛОВО	5510

2. Histogram and kernel density of variable Distance and log(Distance) were built via tools of library SeaBorn which are shown below.

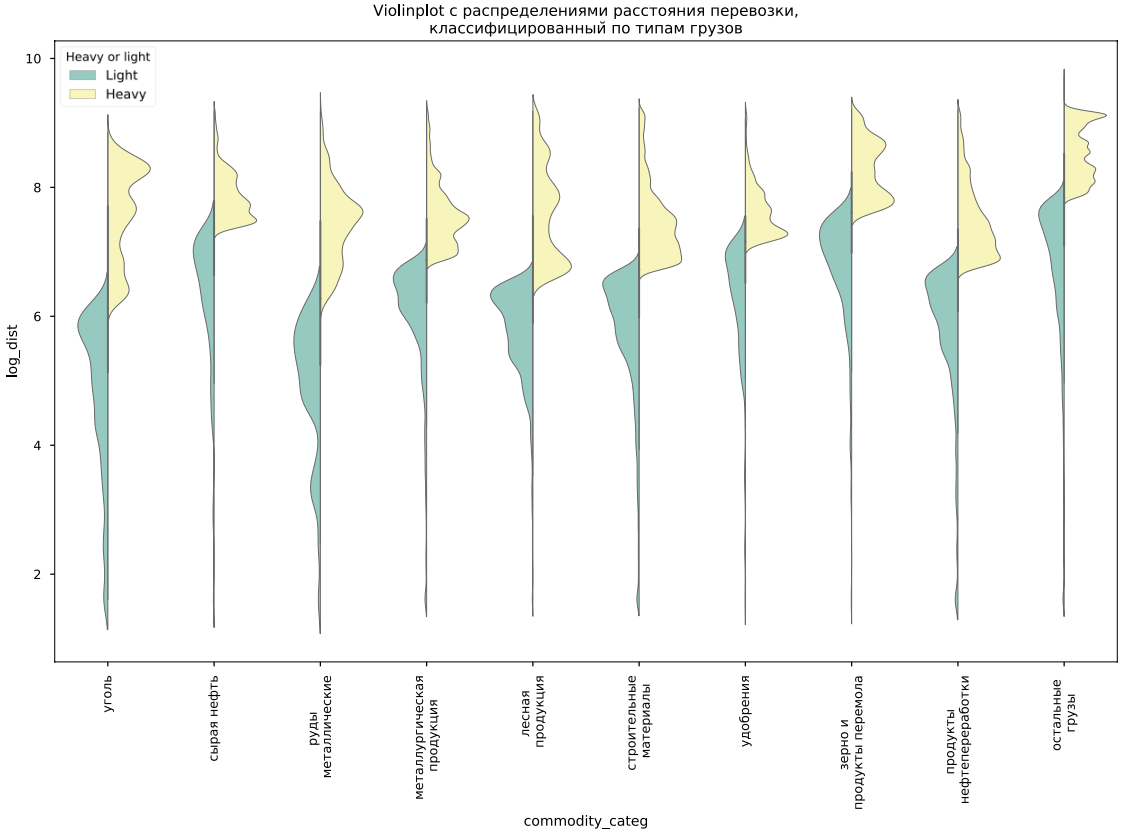




3. Boxplot of log(Distance) is presented below. The biggest spread has Coal but the lowest Oil and Fertilizes. The grains is transported on the most far distances.



4. Violin plot for each type of commodity and specified by the weight is shown below.



5. To draw scatter plot where ordinate is log(amount) we need to decide what to do with 0 in the ‘amount’ column. Fortunately, there are only 23 rows where ‘amount’ = 0. Taking into consideration the fact that number of observations is 257 483, I have dropped these rows. Also there are NaN values in the column ‘amount’ which I have dropped too because their amount is relatively small, 1696. The amount of left data is 255 787.