

УДК [378.147:314]:004

Панченко Любов Феліксівна

професор, доктор педагогічних наук, професор кафедри соціології
НТУУ «Київський політехнічний інститут ім. І.Сікорського», м. Київ, Україна
ORCID ID 0000-0002-9979-0625
lubov.felixovna@gmail.com

ПІДГОТОВКА МАЙБУТНІХ СОЦІОЛОГІВ ДО КОМП'ЮТЕРНОГО АНАЛІЗУ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ І СТРУКТУР

Анотація. У сучасному цифровому глобалізованому світі зростає важливість підготовки майбутніх соціологів у галузі демографії й демографічної статистики, яка основана не лише на демографічних теоріях, а й практичному застосуванні нових комп'ютерних засобів і технологій, баз даних і сервісів Інтернету. У статті аналізуються можливості сучасних комп'ютерних засобів щодо аналізу демографічних процесів і структур з метою навчання майбутніх соціологів; обґрунтовується використання середовища R як інструмента аналізу і графічного представлення демографічних даних; презентується ідея навчання студентів комп'ютерному аналізу даних демографічних досліджень шляхом поєднання використання електронних таблиць Excel, статистичного пакету SPSS, середовища R, яка ілюструється двома прикладами. Перший приклад стосується побудови і порівняння статево-вікової піраміди населення України за різні роки і включає пошук відповідних даних, побудову піраміди стандартними засобами для побудови діаграм Excel; засобом SPSS (ChartBuilder, Histogram, PopulationPyramid), з використанням пакету rpyramid середовища R. Другий приклад, присвячений розрахункам коефіцієнтів демографічного навантаження дітьми і прабатьками, візуалізації їх динаміки, передбачає знайомство з демографічним паспортом населення України. У статті представлено розроблене методичне забезпечення з навчання студентів-соціологів аналізу демографічних даних, яке містить лекції-презентації з основ роботи в середовищі R та R Studio; лабораторні роботи (коротка теорія, детальні інструкції до виконання, контрольні запитання, завдання до самостійної роботи); масиви даних, які додаються до кожної роботи. Аналізуються дидактичні можливості та шляхи використання вільного сервісу Garminder, який включає перелік інструментів під загальною назвою «Playwithdata»: бульбашкову діаграму, карти, ранжування, тренди, вікові піраміди; усі вони дозволяють отримати яскраві динамічні візуалізації обраних демографічних показників (залежно від дослідницького питання) за країнами й континентами в часі, що мотивує студентів на додатковий науковий пошук.

Ключові слова: комп'ютерний аналіз даних; демографічна статистика; Excel; SPSS; середовище R; інформаційно-комунікаційні технології; Garminder; студенти-соціологи.

1. ВСТУП

Постановка проблеми. Невід'ємною частиною підготовки майбутніх соціологів є осмислення суті демографічних явищ і проблем, а також оволодіння методами вимірювання й аналізу демографічних процесів і структур. Наряду з вивченням історії становлення демографії як науки, студенти-соціологи з'ясовують зміст джерел демографічної інформації; опановують методи дослідження динаміки чисельності населення, його структури, окремих демографічних процесів; оцінюють та виявляють закономірності демографічних процесів, тощо. Наразі зростає важливість підготовки майбутніх соціологів у галузі демографії та демографічної статистики, яка основана не лише на демографічних теоріях, а й практичному застосуванні нових інструментів дослідження демографічних явищ – комп'ютерних засобів і технологій, баз даних демографічної інформації, сервісів Інтернету.

Аналіз останніх публікацій. Основи демографії і демографічної статистики представлені у роботах вітчизняних науковців С. Стеценко [2], З. Пальян [3], І. Прибуткової [4] та ін.

Демографію як просторову соціальну науку, яка має справу з розподіленими в просторі даними, розглядає у своєму аналітичному огляді П. Восс (P. Voss) [1]. Прикладний аналіз просторових даних (Exploratory Spatial Data Analysis) включає теоретичні й методологічні питання аналізу просторово розподілених даних: типи даних, які доступні для просторового аналізу; основні способи їх представлення та аналізу; просторові закономірності.

Останнім часом цікавість науковців різного фаху привертає середовище R, з його різноманітністю інструментів для статистичного аналізу й візуалізації даних у різних галузях. Розвиток середовища R, методичні аспекти його застосування представлені у роботах Р. Андерсена (R. Andersen), Дж. Фокса (J. Fox) [5] та ін. Р.С. Біванд (R.S. Bivand), Е. Пезезма (E. Pebesma) та В. Гомез-Рубіо (V. Gómez-Rubio) [6] розглядають прикладний аналіз просторових даних за допомогою R. Пакети для обробки демографічних даних, представлення просторової інформації в середовищі R розроблені зусиллями А. Ламштейна (A. Lamstein) [7], Р. Гиндмана (R. Hyndman) [8], Зака У. Алмквіста (Zack W. Almquist), Юнь Тайн (YunTai), Т. Ріффе (T. Riffe) та ін. [9]. Роботи С. Хауба (C. Haub) присвячені обробці даних переписів США [10]. Використання R для навчання обробки демографічних даних розглядається в роботах К. Камарда (C. Camarda) [11]. Зауважимо, що більшість розробок представлено переважно у вигляді технічної документації до пакетів R, що потребує методичної обробки для навчання студентів-гуманітаріїв.

На жаль, середовище R не посіло належного місця для навчання студентів обробки демографічних даних. Ми не знайшли такого курсу у програмі вітчизняних вишів, натомість знайшли короткий курс з програмування в R для завдань демографії в Іспанії, у рамках якого для студентів PhD дослідників Spanish National Research Council розглядаються основні статистики і графіки; розподіли, випадкові числа та ручна оптимізація; генералізовані лінійні моделі, демографічне прогнозування, матрична алгебра та когортний аналіз (<https://sites.google.com/site/carlogiovannicamarda>). Ще один курс з літньої школи в Барселоні охоплює три сильні сторони R: статистичний аналіз, візуалізацію даних та просторовий аналіз. Перший модуль: «Вступ до R і базової статистики» представляє маніпулювання даними і базовий статистичний аналіз. Другий модуль «Візуалізація даних і просторовий аналіз» зосереджується на інструментах візуалізації даних, що представлені у пакеті ggplot2 R. Далі студенти навчаються обробляти просторові дані в R, розробляти цифрову картографію і виконувати основні види аналізу просторових даних (<http://ced.uab.cat/BarcelonaSummerSchool%207.pdf>).

Отже, можна спостерігати протиріччя між широкими можливостями сучасних комп'ютерних засобів, зокрема середовища R щодо аналізу демографічних процесів та недостатністю використання у курсах з демографії для студентів-соціологів.

Мета статті: проаналізувати можливості сучасних комп'ютерних засобів щодо аналізу демографічних даних з метою навчання майбутніх соціологів, обґрунтувати використання середовища R як одного з інструментів такого аналізу, відібрати необхідні для цього пакети R, презентувати приклади навчання студентів комп'ютерному аналізу демографічних даних шляхом поєднання електронних таблиць, статистичного пакету SPSS, середовища R.

2. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Аналіз наукових джерел [2]-[4] показав, що демографічна статистика використовує арсенал методів загальної теорії статистики, але й власні, оригінальні методи. Стан і склад населення, вивчаються загальними для статистичної науки способами, а вивчення демографічних процесів потребує розвитку системи особливих методів, серед яких можна виділити демографічні таблиці, демографічні сітки тощо. Загальна схема вивчення демографічних процесів складається з таких етапів: екстенсивний аналіз новоутворених за певний час сукупностей, тобто новонароджених, новопомерлих і т. і.; інтенсивний аналіз, мета якого зіставити новоутворені сукупності з існуючим населенням; моделювання кожного процесу; їх синтез під час моделювання населення в цілому [2, с.121].

Отже, під час аналізу даних демографічних досліджень майбутні соціологи мають застосовувати загальні статистичні методи аналізу даних, які вони вже вивчали в рамках курсів з кількісних методів аналізу даних дослідження. Ці методи мають комп'ютерну реалізацію в електронних таблицях Excel у вигляді статистичних функцій, стандартних діаграм, Пакету аналізу. Статистичний пакет SPSS, який є стандартом у галузі обробки даних у соціальних науках, містить описову статистику, кореляційний і регресійний аналіз, широкий перелік методів багатовимірного аналізу тощо, те ж був предметом вивчення в курсах, що передували вивченню демографії. Тому в навчанні студентів-соціологів аналізу демографічних процесів і структур пропонуємо спиратися на ці комп'ютерні середовища, розширивши їх за рахунок вільного середовища R [12].

Середовище R включає функції описової статистики, інструменти для розвідувального аналізу даних, методи багатовимірного аналізу, нейронні мережі, дерева рішень, широкий набір інструментів для візуалізації даних тощо. Функціональність R може розширюватися за рахунок пакетів, які доступні на сайті CRAN, і які можуть бути написані користувачами-фахівцями різних галузей. Середовище R розвивається зараз з участю всієї світової спільноти. Ось список з 162 книг, пов'язаних з R (<http://www.r-project.org/doc/bib/R-books.html>).

Головною перевагою R у даному випадку вважаємо той факт, що в ньому у вигляді пакетів реалізовані специфічні для демографії методи аналізу, зокрема демографічні таблиці, демографічні сітки, демографічні коефіцієнти тощо.

Аналіз наукових джерел [5]-[15], власний досвід викладання дозволяє виділити список пакетів, які доцільно використовувати для навчання студентів обробки даних демографічної статистики (табл.1).

Таблиця 1

Пакети R, які доцільно використовувати для навчання студентів аналізу демографічних даних

Назва пакету	Короткий зміст
demography	Функції для загальних розрахунків, що використовуються в демографії: побудова таблиць смертності, моделювання Лі-Картера; аналіз функціональних даних щодо показників смертності, показників народжуваності, міграцій; стохастичного прогнозування
pyramid	Побудова популяційної піраміди
ggplot2	Візуалізація, інтерактивні графіки
choroplethr, choroplethrMaps	Фонова картограма, хороплет (від грец. χόρο– + πλῆθ[ος], «місце, область» + «значення») – вид картограми, на якій штриховкою різної густоти або фарбою різного ступеня насиченості

	зображують інтенсивність будь-якого показника в межах територіальної одиниці
acs	Пошук таблиць у даних переписів у США. Завантаження, опрацювання даних American Community Survey та даних переписів
census, census2010	Опрацювання даних останніх двох переписів США (2000 р. та 2010 р.)
LexisPlotR	Сітка Лексиса (графічне зображення сукупностей людей і подій у їх житті)
RJSONIO	Перетворення даних різних типів
censusapi	Використання даних переписів США, отримання власного API для доступу до бази даних переписів США
sp	Креслення просторових полігонів для даних GADM
ggmap	Просторові карти

У нагоді для викладача також стануть деякі вбудовані в R масиви демографічних даних різних країн (табл.2).

Таблиця 2

Вбудовані масиви демографічних даних

Назва масиву	Зміст
uspor	Популяція США за даними переписів з 1790–1970 рр.
fr.mort	База даних смертності Франції
aus.ft	База даних фертильності жінок Австралії

Дані демографічної статистики по різних країнах можна знайти в світових базах смертності, плідності, даних Держкомстату України, європейської статистики.

Розглянемо в рамках цієї статті два приклади лабораторних робіт з аналізу демографічних структур і процесів.

Перший приклад стосується побудови статеві-вікової піраміди населення України. Як відомо, віково-статеві піраміди є діаграмами, на яких кількість людей кожного віку або їх частка у населенні у відсотках зображується горизонтальною смугою певного масштабу. Смуги розташовуються одна над іншою в порядку збільшення значень віку, у лівій частині діаграми – для чоловіків, у правій – для жінок. Такі піраміди будуються, зазвичай, за п'ятирічними віковими групами, і дозволяють аналізувати типи відтворення населення у різних країнах і регіонах, або порівнювати дані однієї країни в часі.

Порівняємо дані переписів 1959р. та 2001 р. в Україні.

Спочатку знайдемо інформацію про перепис 1959 р. в Україні.

1. Скористаємося адресою бази даних переписів України:
http://database.ukrcensus.gov.ua/MULT/Database/Census/databasetree_uk.asp

2. Оберемо «Статеві-віковий склад населення, Розподіл наявного населення за статтю і віковими групами», що відображено на рис.1.

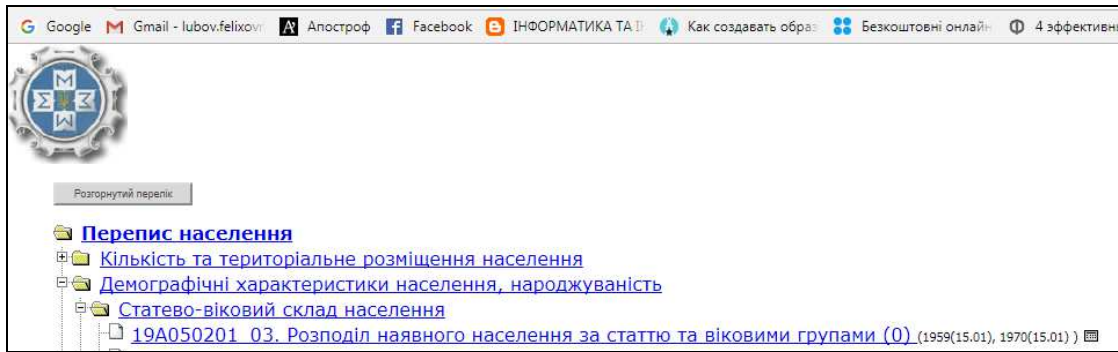


Рис.1. Пошук у базі даних переписів України

3. У відповідній таблиці 19A050201_03 виберемо дані перепису 1959 р. згідно з рис.2 та натиснемо «Продовжити».

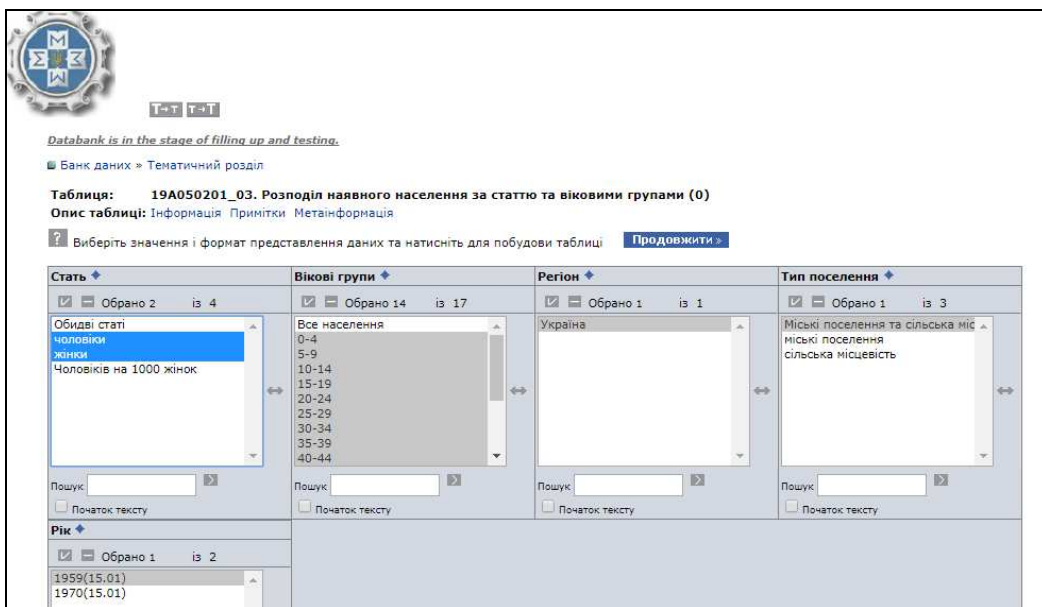


Рис.2. Обирання даних перепису 1959 р. для чоловіків і жінок за віковими групами

4. Отримаємо дані перепису 1959 року в такому вигляді, як показано на рис. 3.

Розподіл наявного населення за статтю та віковими групами (осіб) - Стать, Вікові групи, Регіон, Тип поселення , Рік	
	Україна
	Міські поселення та сільська місцевість
	1959(15.01)
чоловіки	
0-4	2 040 066
5-9	1 989 679
10-14	1 511 731
15-19	1 742 170
20-24	2 012 115
25-29	1 611 944

Рис.3. Фрагмент даних про чисельність чоловіків за віковими групами за переписом 1959 р.

5. Збережемо це як файл Excel і побудуємо за цими даними популяційну піраміду в Excel або SPSS (для побудови окремо надаються детальні інструкції).

На наступному кроці студенти знаходять дані останнього перепису в Україні, який відбувався у 2001 р. (<http://2001.ukrcensus.gov.ua/publications/#p2>), та порівнюють статево-вікову піраміду населення 1959 та 2001р.

Розглянемо далі два способи побудови популяційної піраміди засобами пакету *pyramid* середовищі R.

1 спосіб. Для побудови піраміди ми маємо задати вікові інтервали осіб, а також відсоток чоловіків і жінок для цих вікових інтервалів. Отже, встановивши попередньо пакет *pyramid*, ми можемо написати таку послідовність команд:

```
## Задати вікові інтервали:
ages<-c('0-4', '5-9', '10-14', '15-19', '20-24', '25-29', '30-34', '35-39', '40-44', '45-49', '50-54', '55-59', '60-64', '65-69', '70+')

## Задати відсотокчоловіків в інтервалах:
males<- c(10.98, 10.71, 8.14, 9.38, 10.83, 8.68, 9.05, 5.39, 4.38, 5.62, 4.94, 3.67, 2.96, 2.23, 3.04)

## Задати відсотокжінок в інтервалах:
females<- c(8.36, 8.22, 6.3,7.59, 9, 7.47, 9.22, 6.81, 5.82, 7.35, 6.16, 5.43, 4.09, 3.23, 4.95)

## Об'єднати дані в один датафрейм:
data<- data.frame(males,females,ages)

##Побудувати піраміду
library(pyramid)
pyramid(data)
```

2 спосіб. Попередньо готуємо дані в електронних таблицях Excel і зберігаємо як *csv-файл*ukraine1959R.csv (рис.4).

	A	B	C
1	male	female	age
2	10,98	8,36	0-4'
3	10,71	8,22	5-9'
4	8,14	6,30	10-14'
5	9,38	7,59	15-19'
6	10,83	9,00	20-24'
7	8,68	7,47	25-29'
8	9,05	9,22	30-34'
9	5,39	6,81	35-39'
10	4,38	5,82	40-44'
11	5,62	7,35	45-49'
12	4,94	6,16	50-54'
13	3,67	5,43	55-59'
14	2,96	4,09	60-64'
15	2,23	3,23	65-69'
16	3,04	4,95	70 +'

Рис.4. Дані перепису України 1959 р., підготовлені в електронних таблицях як *csv-файл*

У середовищі R використовуємо команди:

```
## Читаємо файл з даними:
data1959<-read.csv("ukraine19591R.csv", header = T, sep = ";",dec = ",")

## Будуємо популяційну піраміду
library(pyramid)
pyramid(data1959)
```

Аналізуємо результати, наведені на рис.5.

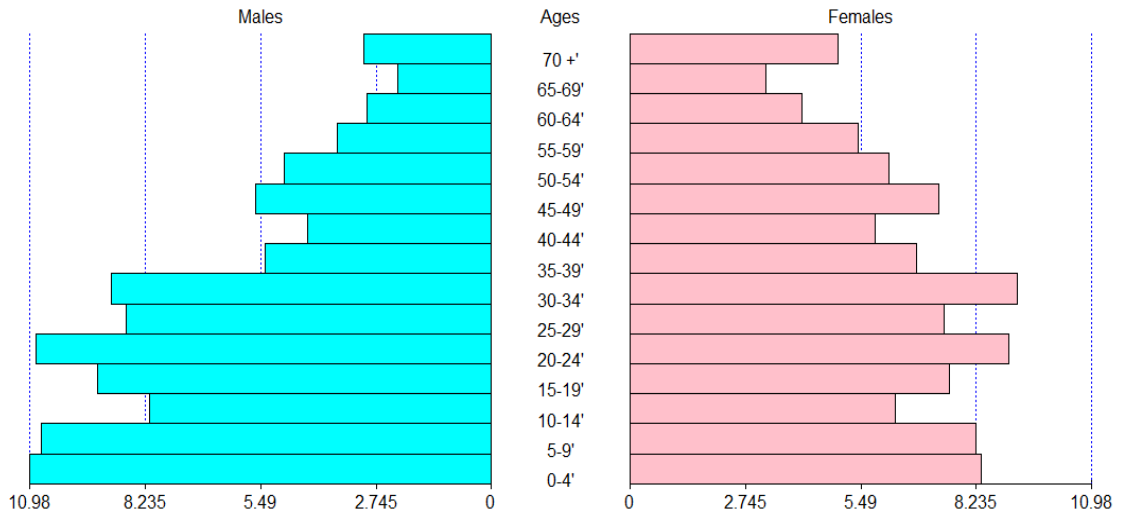


Рис.5. Популяційна піраміда населення України за даними перепису 1959 р. (у %) [16], побудована в R

Зауважимо, що в ході розробленого нами практикуму студенти також будують такі піраміди за допомогою стандартних діаграм в Excel і конструктора діаграм пакету SPSS.

Другий приклад призначений для ілюстрації розрахунків коефіцієнтів демографічного навантаження. Мета цієї роботи: надати досвід розрахування коефіцієнтів демографічного навантаження на прикладі України в середовищах Excel, SPSS, R.

У ході цієї роботи студенти знайомляться з демографічним паспортом населення України (рис. 6), (http://database.ukrcensus.gov.ua/MULT/Dialog/statfile_c_files/pasport.files/pasport/00_uk.htm#0202).

Регіон	Україна									
Рік	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Міські поселення та сільська місцевість										
У віці 0-14 років										
Обидві статі	6501091	6476188	6483560	6495990	6531531	6620598	6710689	6449171	6494293	6535536
чоловіки	3336144	3325162	3330609	3338373	3358132	3405479	3453718	3319634	3343782	3365550
жінки	3164947	3151026	3152951	3157617	3173399	3215119	3256971	3129537	3150511	3169986

У віці 15-64 роки										
Обидві статі	3218451 4	3216979 5	32130170	3213696 8	3199331 1	3184677 6	3160637 4	2963471 0	2932772 4	2901183 5
чоловіки	1543031 2	1541000 3	15390314	1539087 5	1533029 4	1527711 2	1517225 9	1424974 9	1411674 5	1398075 0
жінки	1675420 2	1675979 2	16739856	1674609 3	1666301 7	1656966 4	1643411 5	1538496 1	1521097 9	1503108 5
У віці 65 років і старше										
Обидві статі	7506704	7317376	7168862	6965221	6928440	6905318	6928831	6675780	6768862	6867534
чоловіки	2531222	2449767	2386144	2303368	2288286	2280153	2292311	2218443	2257354	2298280
жінки	4975482	4867609	4782718	4661853	4640154	4625165	4636520	4457337	4511508	4569254

Рис. 6. Розподіл постійного населення України за основними віковими групами (осіб), на початок року

Далі обчислюється демографічне навантаження на покоління батьків в електронних таблицях і представляється так (рис. 7).

Регіон	Україна									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Загальне навантаження	435	429	425	419	421	425	432	443	452	462
Навантаження особами у віці 0-14 років	202	201	202	202	204	208	212	218	221	225
Навантаження особами у віці 65 років і старше	233	228	223	217	217	217	220	225	231	237

Рис. 7. Розраховані коефіцієнти загального навантаження, навантаження дітьми, прабатьками

Результати обчислення візуалізуються у вигляді графіків, які ілюструють зміну демографічного навантаження з часом (рис.8).



Рис.8. Візуалізація загального демографічного навантаження, Україна

Наступним етапом є представлення тих самих даних в SPSSта проведення аналогічних розрахунків і побудова графіків. Для цього студенти мають транспонувати дані так, як подано в табл. 3.

Таблиця 3

Транспоновані дані для обчислення коефіцієнтів навантаження в SPSS

Year	t0_14	m0_14	f0_14	t15_64	m15_64	f15_64	t65	m65	f65
2008	6501091	3336144	3164947	32184514	15430312	16754202	7506704	2531222	4975482
2009	6476188	3325162	3151026	32169795	15410003	16759792	7317376	2449767	4867609
2010	6483560	3330609	3152951	32130170	15390314	16739856	7168862	2386144	4782718
2011	6495990	3338373	3157617	32136968	15390875	16746093	6965221	2303368	4661853
2012	6531531	3358132	3173399	31993311	15330294	16663017	6928440	2288286	4640154
2013	6620598	3405479	3215119	31846776	15277112	16569664	6905318	2280153	4625165
2014	6710689	3453718	3256971	31606374	15172259	16434115	6928831	2292311	4636520
2015	6449171	3319634	3129537	29634710	14249749	15384961	6675780	2218443	4457337
2016	6494293	3343782	3150511	29327724	14116745	15210979	6768862	2257354	4511508
2017	6535536	3365550	3169986	29011835	13980750	15031085	6867534	2298280	4569254

Для створення відповідного файлу в SPSS змінні треба описати (рис.9), перенести відповідні дані.

	Name	Type	
1	Year	Numeric	Рік
2	t0_14	Numeric	Населення віком 0-14 років
3	m0_14	Numeric	Чоловіки віком 0-14 років
4	f0_14	Numeric	Жінки 0-14 років
5	t15_64	Numeric	Населення віком 15-64 роки
6	m15_64	Numeric	Чоловіки 0-14 років
7	f15_64	Numeric	Жінки 0-14 років
8	t65	Numeric	Населення старше 65 років
9	m65	Numeric	Чоловіки 65+
10	f65	Numeric	Жінки 65+.
11	nt	Numeric	Коефіцієнт загального навантаження
12	n0_14	Numeric	Коефіцієнт навантаження дітьми
13	n65	Numeric	Коефіцієнт навантаження прабатьками

Рис. 9. Опис змінних у SPSS

Далі можна розрахувати загальний коефіцієнт навантаження за допомогою меню Transform, Compute Variable(рис.10).

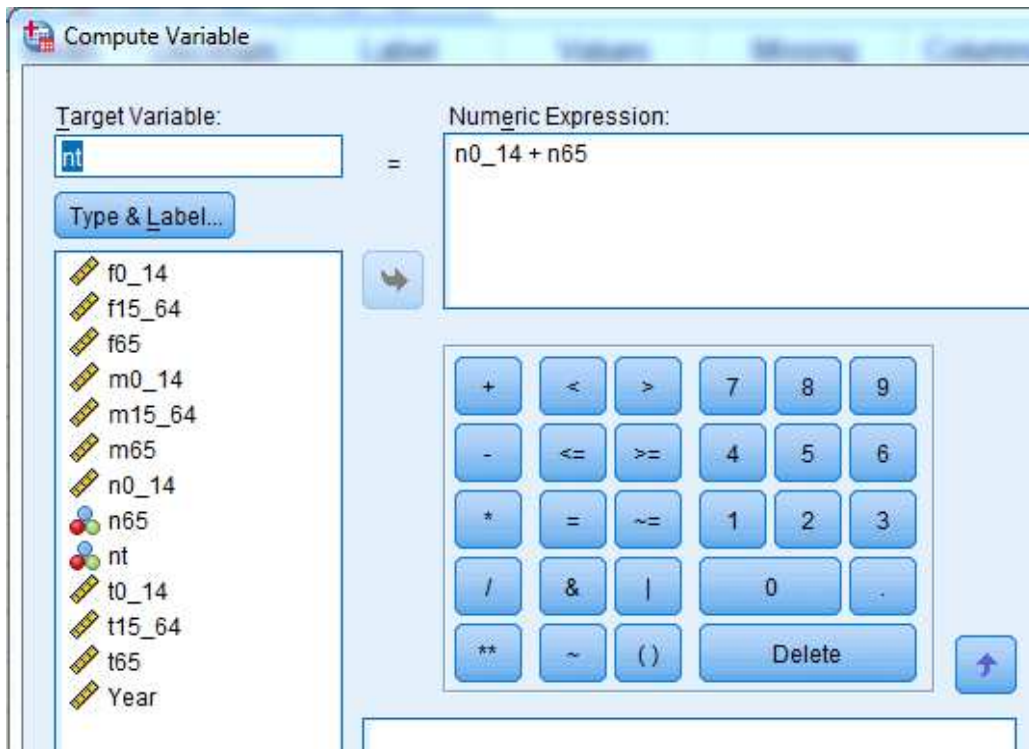
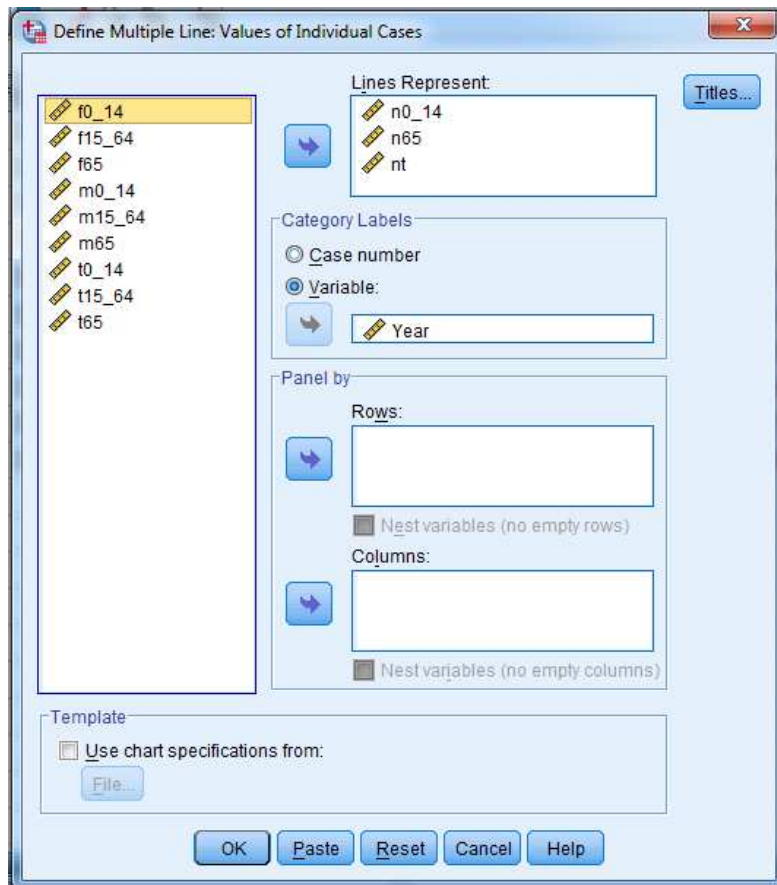


Рис.10. Розрахунки коефіцієнтів навантаження

Для побудови графіка використовуються Graphs, LegacyDialogs (рис. 11).



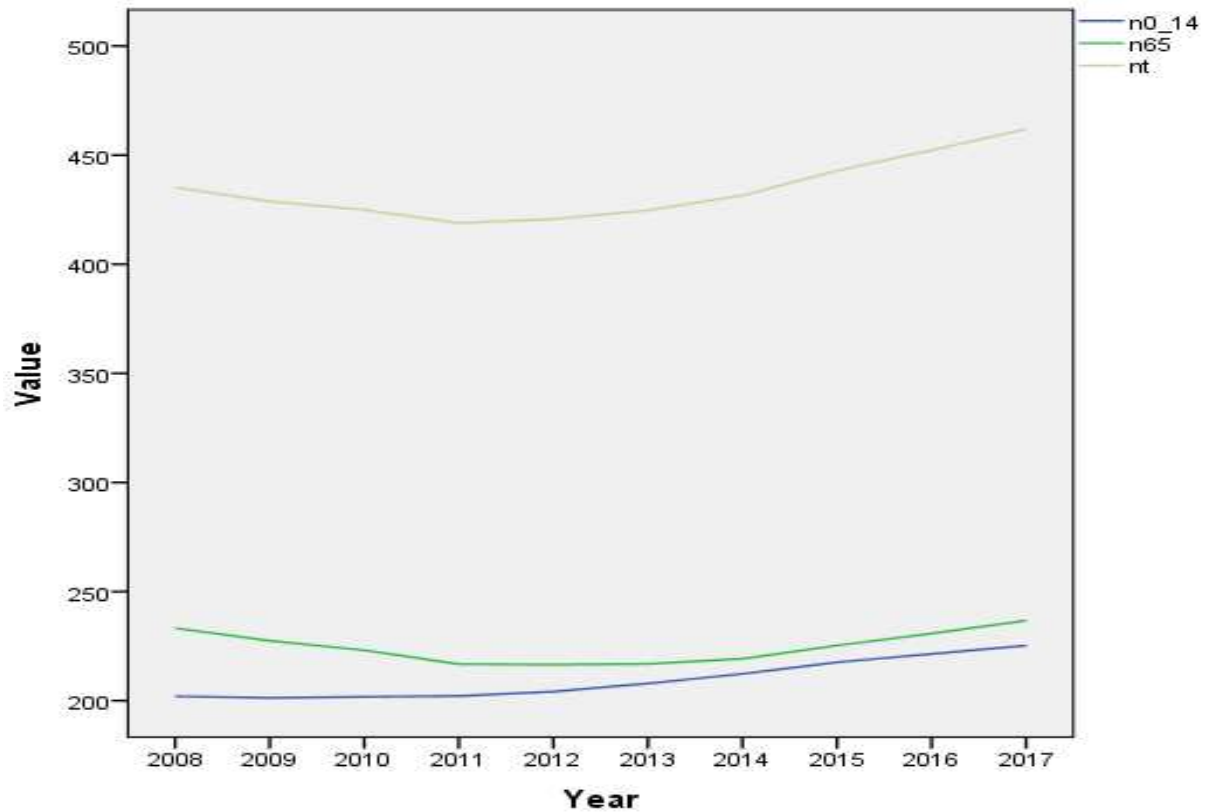


Рис.11. Побудова графіка в SPSS

Наприкінці цього етапу студенти зберігають файл у форматі csv з ім'ям nUkraineInput.csv для подальшої роботи в середовищі R.

У середовищі R спочатку треба прочитати файл з даними (він має бути в поточному каталозі R)

```
data1<- read.csv("nUkraineInput.csv", header = T, sep = ";",dec = ",", )
```

Щоб з'ясувати структуру даних, доцільно переглянути перші 3 рядки :

```
head(data1, 3)
```

```
Year t0_14 m0_14 f0_14 t15_64 m15_64 f15_64 t65 m65 f65
1 2008 6501091 3336144 3164947 32184514 15430312 16754202 7506704 2531222 4975482
2 2009 6476188 3325162 3151026 32169795 15410003 16759792 7317376 2449767 4867609
3 2010 6483560 3330609 3152951 32130170 15390314 16739856 7168862 2386144 4782718
  n0_14n65nt
1 201.9944 233.2396 435.2340
2 201.3127 227.4611 428.7738
3 201.7904 223.1193 424.9097
```

А потім обчислити коефіцієнти навантаження такими командами:

```
data1$n0_14 <- data1$t0_14 / data1$t15_64*1000
```

```
data1$n65 <- data1$t65 / data1$t15_64*1000
```

```
data1$nt<- data1$n0_14+data1$n65
```

Для побудови графіку базовими засобами R використовується функція plot (рис. 12).



Рис. 12. Графік загального навантаження, побудований у R

Як узагальнююче завдання студентам пропонується порівняти отримані дані з класифікацією Г. Зунберга [2, с.51] (табл. 4). У рамках цього завдання вони мають використовувати статистичні критерії для порівняння розподілів, зокрема критерій Пірсонах 2.

Таблиця 4

Класифікація вікової структури населення за Г. Зунбергом

Вікова група	Вікова структура населення, %		
	прогресивна	стаціонарна	регресивна
Діти	40	27	20
Батьки	50	50	50
Прабатьки	10	23	30
Разом	100	100	100

Розроблене методичне забезпечення з навчання студентів-соціологів аналізу демографічних даних містить лекції-презентації з основ роботи в середовищі R та R Studio; лабораторні роботи (коротка теорія, детальні інструкції до виконання, контрольні запитання, завдання до самостійної роботи); масиви даних, які додаються до кожної роботи. Більшість робіт побудована так, що студенти знаходять дані, обчислюють демографічні коефіцієнти в Excel, SPSS, R та візуалізують отримані результати за допомогою різного роду графіків, а потім їх порівнюють й інтерпретують.

Теми лабораторних робіт включають такі:

- Побудова популяційної піраміди засобами Excel, SPSS, R.
- Аналіз демографічних даних засобами R.
- Дослідження даних перепису США в середовищі R.
- Обчислення коефіцієнтів навантаження прабатьками та дітьми.
- Дослідження структури сімей різних країн.
- Національний та етнічний склад населення.
- Статистика зайнятості та безробіття.
- Розрахунки в таблицях дожиття.
- Дослідження даних бази смертності.
- Використання пакету demography середовища R.
- Побудова сітки Лексіса засобами R.

- Розрахунок показників плідності. Аналіз даних міжнародних баз даних плідності.
- Статистика шлюбності та розлучуваності.
- Обчислення та візуалізація чистої та валової міграції (Excel, SPSS, R.).

У ході робіт студенти реєструються в світових базах смертності та плідності, отримують власне API для доступу до даних переписів США, працюють з даними статистики населення України.

Як заключне заняття пропонуємо студентам-соціологам роботу з вільним сервісом Garminder (рис. 13). Garminder–незалежна шведська установа, яка випускає безкоштовні навчальні ресурси, спрямовані на покращання розуміння світу на основі надійної статистики; вона співпрацює з університетами, ООН, державними установами та неурядовими організаціями. Метою фундації Garminder є «сприяння сталому глобальному розвитку та досягнення Цілей тисячоліття ООН зі збільшення використання і розуміння статистичних даних та іншої інформації про соціальний, економічний та екологічний розвиток на місцевому, національному та глобальному рівнях» [17].

Сервіс Garminder (рис. 14) включає набір інструментів під загальною назвою «Playwithdata» бульбашкову діаграму, карти, ранжування, тренди, вікові піраміди; усі вони дозволяють отримати динамічну візуалізацію обраного демографічного показника (залежно від дослідницького питання) за країнами та континентами в часі (від 1800 до 2015 р.). Так, для категорії «освіта» можна порівнювати академічні досягнення, грамотність, гендерну рівність, середню кількість років навчання, відсоток тих, хто закінчив школу і т. і.

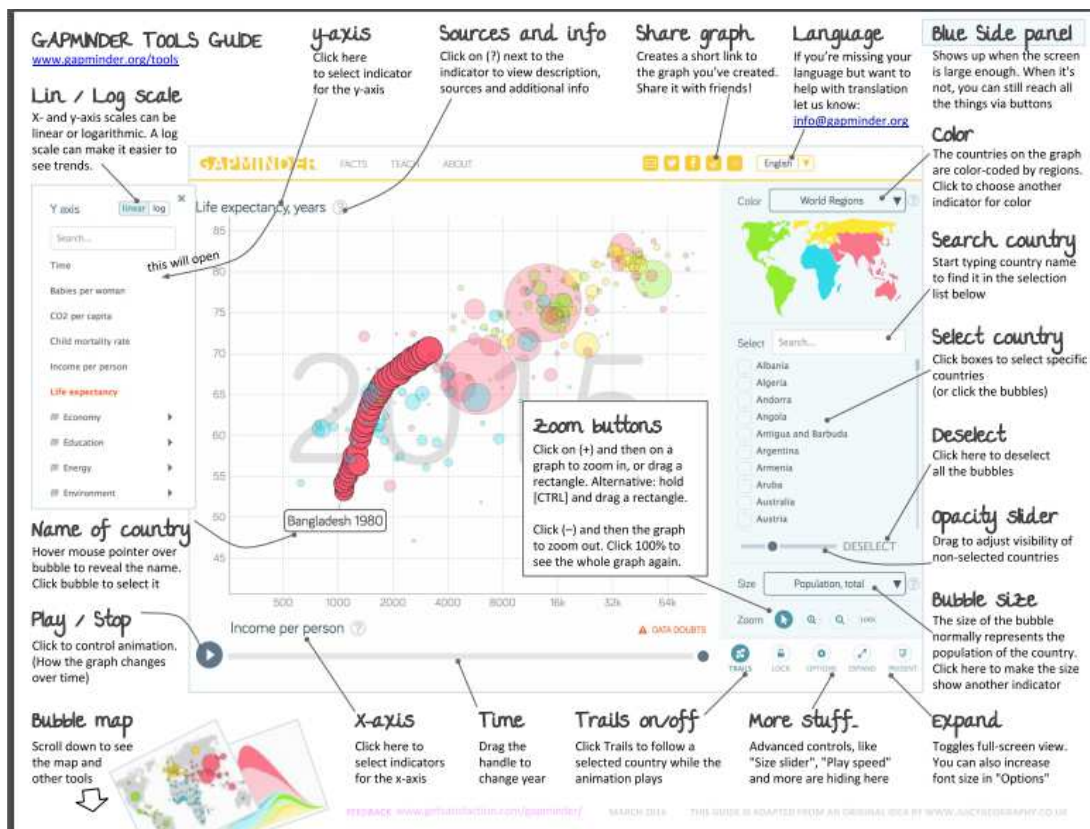


Рис.13. Інструкція для роботи із сервісом Garminder

Зауважимо, що сервіс має дружній інтерфейс, робота з ним викликала зацікавленість студентів, які будували популяційні піраміди, порівнювали демографічні показники різних країн у динаміці, створювали рейтинги країн і регіонів за різними показниками, ставили власні дослідницькі запитання й відшукували на них відповідь, вивчали *DollarStreet*, інтерактивне відображення світу як вулиці, де номер на вулиці відповідає щоденному доходу на людину в сім'ї.



Рис. 14. Інструменти *Gapminder*: бульбашкові діаграми, карти, ранжування, тренди, вікові піраміди

Часткова відповідь на дослідницьке запитання: «Який тип відтворення населення притаманний різним континентам світу?» представлена на рис.15. Вона стосується Африки й отримана за допомогою інструмента *AGES* (вікова піраміда).

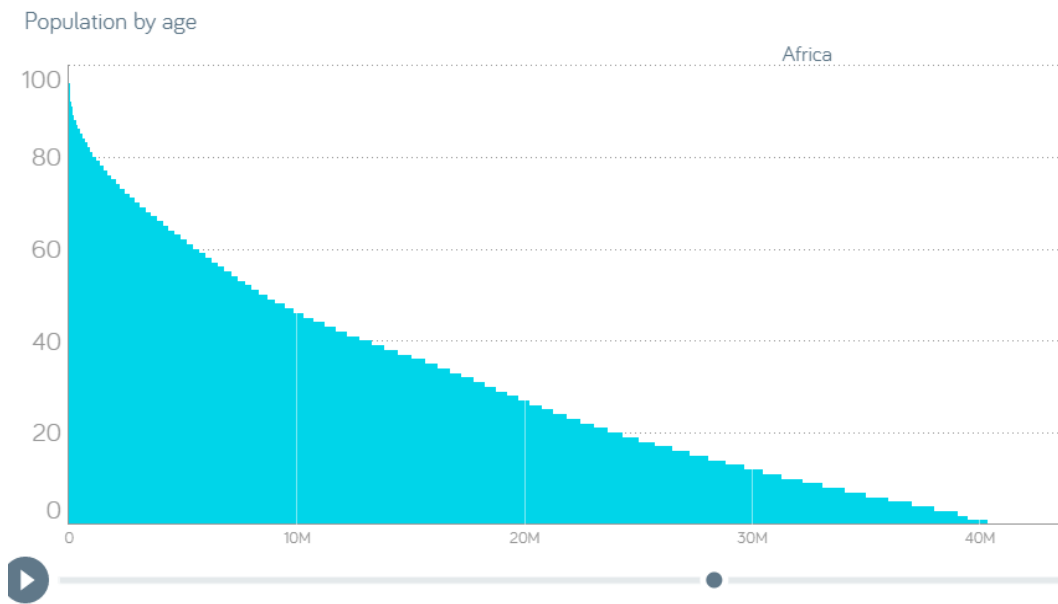


Рис. 15. Вікова піраміда населення Африки, побудована за допомогою інструмента «*AGES*» *Gapminder*

На заняттях і в самостійній роботі можна використовувати і низку інших проектів, які пропонує *Gapminder*, а саме: *Всесвітній розподіл доходів*, що містить інтерактивне відображення статистичних даних про розподіл доходів домашніх господарств у Бангладеш, Бразилії, Індії, Індонезії, Китаї, Нігерії, Пакистану, США, Японії і світу (1970-1998); *Тенденції розвитку людського потенціалу 2003*, тематичну анімовану презентацію для Доповіді про розвиток людства 200; *Всесвітню діаграму здоров'я 2001*, яка відображає розвиток охорони здоров'я для всіх країн світу з часовими рядами для 35 показників.

3. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У сучасному глобалізованому цифровому світі зростає важливість підготовки майбутніх соціологів у галузі демографії й демографічної статистики із застосуванням нових комп'ютерних засобів і технологій. З'ясовано, що демографічна статистика використовує загальний арсенал статистичних методів для вивчення складу і стану населення та специфічні методи для вивчення демографічних процесів, зокрема демографічні таблиці, демографічні сітки тощо. Отже, у ході навчання майбутніх соціологів комп'ютерному аналізу демографічних процесів і структур доцільно спиратися на їх попередню підготовку щодо кількісних методів і знайомство зі статистичним пакетом SPSS та електронними таблицями Excel. Пропонується розширити цей перелік за рахунок середовища R, у якому реалізовані не лише загальні статистичні методи, а й специфічні, саме для демографії. Як свідчить наш досвід, можливості середовища R дозволяють з успіхом застосовувати його для навчання студентів аналізу демографічних даних. Серед пакетів R, які доцільно використовувати для обробки демографічної інформації, *d emography*, *piramid*, *ggplot2*, *choroplethr*, *choroplethrMaps*, *acs*, *census*, *census2010*, *LexisPlotR*, *censusapi* та ін. Зауважимо, що демографічні пакети працюють лише з останніми версіями R і RStudio, та з версією Windows 7 та вище. Оболонка R Studio полегшує процес аналізу даних у R. Використання декількох інструментів – Excel, SPSS, R – уможливило порівняння результатів аналізу тих самих даних та їх візуалізації у різних середовищах з метою поглибленого розуміння студентами процесів, що вивчаються. Урізноманітнювати процес навчання аналізу демографічних процесів та мотивувати студентів на самостійний пошук допоможе вільно поширюване програмне забезпечення *Garminder*, що дозволяє формувати дослідницькі питання з демографічного аналізу й відшукувати на них візуалізовану відповідь як онлайн, так і офлайн, порівнюючи в часі країни і континенти.

Напрямки подальшого дослідження ми пов'язуємо з розробкою навчально-методичного забезпечення з демографічного прогнозування в середовищі R, поглибленого вивчення побудови просторових карт і створення звітів з досліджень засобами R Markdown.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] P. V. Voss. «Demography as a Spatial Social Science». *Population Research and Policy Review*, №26, р.457-476, 2007.
- [2] С. Г. Стеценко, *Демографічна статистика*, К., Вища школа, 2005.
- [3] І. М. Прибиткова, *Основи демографії*, К., «АртЕк», 1995.
- [4] З.О. Пальян, *Демографічна статистика*, К., КНЕУ, 2003.
- [5] J. Fox, R.Andersen. «Using the r statistical computing environment to teach social statistics courses». [Електронний ресурс]. Режим доступу: https://r-forge.r-project.org/R/?group_id=328.
- [6] R. S. Bivand, E. Pebesma, V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, New York, 2013.
- [7] A. Lamstein. Exploring the World with Choroplethr [Електронний ресурс]. Режим доступу: https://www.slideshare.net/main/private_slideshow?path_to_redirect_at=https%3A%2F%2Fwww.slideshare.net%2Farilamstein%2Fexploring-the-world-with-choroplethr-febru.
- [8] R. Hyndman «Forecasting time series using R». [Електронний ресурс]. Режим доступу: <https://robjhyndman.com/seminars/melbournrug/>.
- [9] Zack W. Almquist. «US Census Spatial and Demographic Data in R: The USCensus2000 Suite of Packages», *Journal of Statistical Software*, November 2010, Vol. 37, Issue 6. [Електронний ресурс]. Режим доступу: <http://www.jstatsoft.org256>.
- [10] US census 2010. [Електронний ресурс]. Режим доступу: <http://r.adu.org.za/web/packages/UScensus2010/UScensus2010.pdf>.

- [11] С. G. Camarda [Електронний ресурс]. Режим доступу: <https://sites.google.com/site/carlogiovannicamarda>.
- [12] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3- 900051-07-0 [Електронний ресурс]. Режим доступу: <http://www.R-project.org>.
- [13] R-project.org. (2017). R: The R Project for Statistical Computing [Електронний ресурс]. Режим доступу: <http://www.R-project.org>.
- [14] Cran.r-project.org, 2017 [Електронний ресурс]. Режим доступу: <https://cran.r-project.org/web/packages/demography/demography.pdf>.
- [15] Cran.r-project.org, 2017 [Електронний ресурс]. Режим доступу: <https://cran.r-project.org/web/packages/LexisPlotR/LexisPlotR.pdf>.
- [16] Cran.r-project.org, 2017. [Електронний ресурс]. Режим доступу: <https://cran.r-project.org/web/packages/pyramid/pyramid.pdf>.
- [17] Перепис населення України [Електронний ресурс]. Режим доступу : http://database.ukrcensus.gov.ua/MULT/Database/Census/databasetree_uk.asp.
- [18] Garminder [Електронний ресурс]. Режим доступу: <http://www.garminder.org/>.
- [19] Л. Ф. Панченко, *Практикум по аналізу даних*, Луганск, Изд-во ГУ «ЛНУ имени Тараса Шевченко», 2013.

Матеріал надійшов до редакції 09.01.2018 р.

ПОДГОТОВКА БУДУЩИХ СОЦИОЛОГОВ К КОМПЬЮТЕРНОМУ АНАЛИЗУ ДЕМОГРАФИЧЕСКИХ ПРОЦЕССОВ И СТРУКТУР

Панченко Любовь Феликсовна

профессор, доктор педагогических наук, профессор кафедры социологии
НТУУ «Киевский политехнический институт им. И.Сикорского», г. Киев, Украина
ORCID ID 0000-0002-9979-0625
lubov.felixovna@gmail.com

Аннотация. В современном глобализованном цифровом мире возрастает важность подготовки будущих социологов в области демографии и демографической статистики, основанной не только на демографических теориях, но и практическом применении новых компьютерных средств и технологий, баз данных и сервисов Интернет. В статье анализируются возможности современных компьютерных средств анализа демографических процессов и структур с целью обучения будущих социологов; обосновывается использование среды R как инструмента анализа и графического представления демографических данных; презентуется идея обучения студентов компьютерному анализу данных демографических исследований путем совместного использования электронных таблиц Excel, статистического пакета SPSS, среды R, которая иллюстрируется двумя примерами. Первый пример касается построения и сравнения половозрастной пирамиды населения Украины за разные годы и включает поиск соответствующих данных, построение пирамиды стандартными средствами для построения диаграмм Excel; средством SPSS (ChartBuilder, Histogram, PopulationPyramid), с использованием пакета pyramid среды R. Второй пример, посвященный расчетам коэффициентов демографической нагрузки детьми и прародителями, визуализации их динамики, предполагает знакомство с демографическим паспортом населения Украины. В статье представлены разработанное методическое обеспечение по обучению студентов-социологов анализу демографических данных, которое содержит лекции-презентации по основам работы в среде R и R Studio; лабораторные работы (короткая теория, подробные инструкции для выполнения, контрольные вопросы, задания для самостоятельной работы); массивы данных, которые предлагаются к каждой работе. Анализируются дидактические возможности и пути использования свободного сервиса Garminder, который включает перечень инструментов под общим названием «Playwithdata» пузырьковую диаграмму, карты, ранжирование, тренды, возрастные пирамиды; все они позволяют получить яркие динамические визуализации избранных демографических показателей (в зависимости от исследовательского вопроса) по странам и континентам во времени, что мотивирует студентов на дополнительный научный поиск.

Ключевые слова: компьютерный анализ данных, демографические данные, Excel, SPSS, среда R, информационно-коммуникационные технологии, Garminder, студенты-социологи.

TRAINING SOCIOLOGY STUDENTS IN COMPUTER ANALYSIS OF DEMOGRAPHIC PROCESSES AND STRUCTURE

Liubov F. Panchenko

Dr. of Pedagogical Sciences, PhD, Professor, Professor of the Department of Sociology
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine
ORCID ID 0000-0002-9979-0625
lubov.felixovna@gmail.com

Abstract. In the modern era of digital globalization, it is becoming more and more important to train sociology students in the field of demographics and demographic statistics based not only on demographic theories but also on the practical application of the new computer tools and technologies, databases and Internet services. The article analyzes the capabilities of modern computer tools for the analysis of demographic processes and structures in training sociology students; substantiates the use of the R environment as a tool for analysis and graphical representation of demographic data. It presents the idea of teaching students to perform computer analysis of demographic data using a combination of Excel spreadsheets, SPSS statistical package, R environment illustrated by two examples. The first example concerns building and comparing the gender-age pyramid of the population of Ukraine at different years and includes searching for the relevant data, building the pyramid using standard diagram building Excel tools, using SPSS tools (Chart Builder, Histogram, Population Pyramid), and using pyramid package of R environment. The second example relates to calculation of childcare and grandparent care load coefficients, visualizing their dynamics, and includes an introduction to the demographic passport of Ukraine. The article presents the developed methodological support for teaching sociology students to perform demographic data analysis, including presentation-lectures on the fundamental principles of work in R and R Studio environment, laboratory works (theory summary, detailed operative instructions, control questions, tasks for students 'independent work'); data packages attached to every assignment. The author has analyzed the didactic capabilities of the free Gapminder service that includes the list of the tools titled 'Play with Data', bubble chart, maps, ranking, trends, age pyramids. This provides colorful and dynamic data visualization for chosen demographic criteria (depending on the research objectives) by countries and continents over time that stimulates the students to conduct additional scientific research.

Keywords: computer data analysis; demographic data; Excel; SPSS; environment R; information and communication technologies; Gapminder; students-sociologists.

REFERENCES (TRANSLATED AND TRANSLITERATED)

- [1] P.V. Voss. «Demography as a Spatial Social Science». *Population Research and Policy Review*, №26, p.457-476, 2007. (in English)
- [2] S.G. Stecenko, Demographic statistics, K., Higher School Press, 2005. (in Ukrainian)
- [3] I.M. Prybytkova, Fundamentals of Demography, K., «ArtEk», 1995. (in Ukrainian)
- [4] Z.O. Paljjan, Demographic statistics, K., KNEU, 2003. (in Ukrainian)
- [5] J. Fox, R.Andersen. «Using the r statistical computing environment to teach social statistics courses». [Online]. Available: https://r-forge.r-project.org/R/?group_id=328(in English)
- [6] R.S.Bivand, E. Pebesma, V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, New York, 2013. (in English)
- [7] A.Lamstein. Exploring the World with Choroplethr [Online]. Available: https://www.slideshare.net/main/private_slideshow?path_to_redirect_at=https%3A%2F%2Fwww.slideshare.net%2Farilamstein%2Fexploring-the-world-with-choroplethr-febru (in English)
- [8] R. Hyndman «Forecasting time series using R». [Online]. Available: <https://robjhyndman.com/seminars/melbournrurg/>(in English)

- [9] Zack W. Almquist. «US Census Spatial and Demographic Data in R: The UScensus2000 Suite of Packages», *Journal of Statistical Software*, November 2010, Vol. 37, Issue 6. [Online]. Available: <http://www.jstatsoft.org/256> (in English)
- [10] US census 2010. [Online]. Available: <http://r.adu.org.za/web/packages/UScensus2010/UScensus2010.pdf> (in English)
- [11] C. G. Camarda. [Online]. Available: <https://sites.google.com/site/carlogiovannicamarda> (in English)
- [12] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3- 900051-07-0, URL <http://www.R-project.org> (in English)
- [13] R-project.org. (2017). R: The R Project for Statistical Computing. [online] Available at: <http://www.R-project.org> (in English)
- [14] Cran.r-project.org, 2017. [Online]. Available: <https://cran.r-project.org/web/packages/demography/demography.pdf> (in English)
- [15] Cran.r-project.org, 2017. [Online]. Available: <https://cran.r-project.org/web/packages/LexisPlotR/LexisPlotR.pdf> (in English)
- [16] Cran.r-project.org, 2017. [Online]. Available: <https://cran.r-project.org/web/packages/pyramid/pyramid.pdf> (in English)
- [17] Ukraine census. [Online]. Available: http://database.ukrcensus.gov.ua/MULT/Database/Census/databasetree_uk.asp (in Ukrainian)
- [18] Gapminder. [Online]. Available: <http://www.gapminder.org/> (in English)
- [19] L.F. Panchenko, Data analysis practicum, Luhansk, Luhansk Taras Shevchenko National University Press, 2013. (in Russian)



This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.