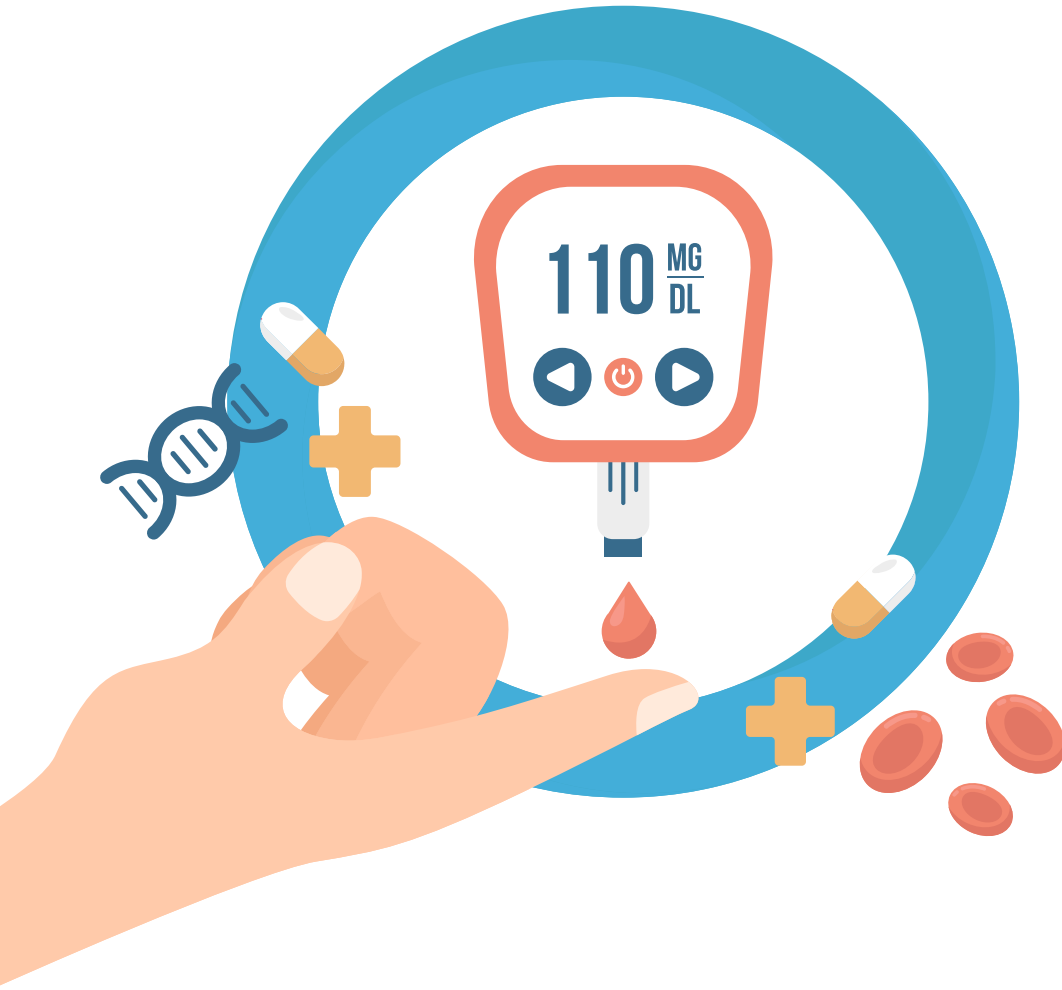


Лабораторна робота №1

Виконали ст. гр. КМ-12
Лазебний О., Якімов Я.,
Петров. В.



- Суть дослідницьких питань піднятих в цій лабораторній роботі полягає в перевірці гіпотез та очікувань щодо наданого набору даних, зокрема, чи є сильним вплив серцево-судинних захворювань, звичок, стану ментального/фізичного здоров'я, рівня освіти, зарплати, віку, статі та використанням медичних послуг на наявність діабету, а також прослідкувати залежності між цими факторами.
- Їх цікавість полягає у дослідженні взаємозв'язків між соціальними, поведінковими та медичними змінними, що дозволить виявити ключові фактори ризику для розвитку діабету. Важливим аспектом є розуміння того, як кожен з цих факторів впливає на ймовірність виникнення діабету, що, у свою чергу, може сприяти розробці ефективних стратегій профілактики цього захворювання.

Опис даних та кроків з їх підготовки

Дані зібрані у рамках щорічного телефонного опитування Центрами контролю та профілактики захворювань США (Centers for Disease Control and Prevention) у 2015 році. Було опитано 441 455 осіб, на основі чого було створено початковий набір даних із 330 змінними: це безпосередні результати опитування кожного та проміжні дані.

В даній роботі використовується набір даних із 22 змінними, створений для побудови класифікаційної моделі для прогнозування того чи є хворою людина на діабет на основі відповідних змінних.



```

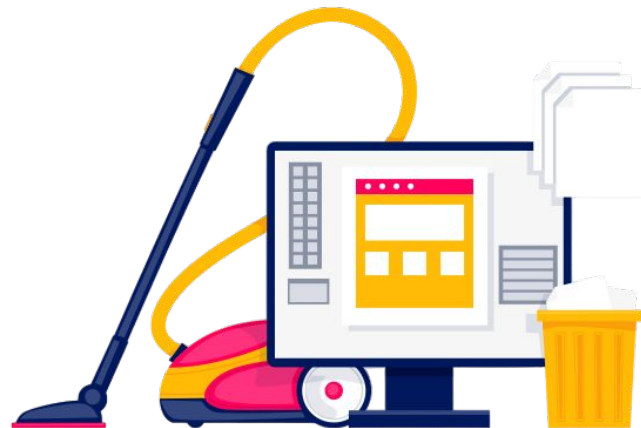
Rows: 253680 Columns: 22
— Column specification —
Delimiter: ",",
dbl (22): Diabetes_binary, HighBP, HighChol, CholCheck, BMI, Smoker, Stro
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this m
spc_tbl_ [253,680 x 22] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Diabetes_binary      : num [1:253680] 0 0 0 0 0 0 0 0 1 0 ...
 $ HighBP               : num [1:253680] 1 0 1 1 1 1 1 1 0 ...
 $ HighChol             : num [1:253680] 1 0 1 0 1 1 0 1 1 0 ...
 $ CholCheck            : num [1:253680] 1 0 1 1 1 1 1 1 1 ...
 $ BMI                  : num [1:253680] 40 25 28 27 24 25 30 25 30 24 ...
 $ Smoker               : num [1:253680] 1 1 0 0 0 1 1 1 1 0 ...
 $ Stroke               : num [1:253680] 0 0 0 0 0 0 0 0 0 0 ...
 $ HeartDiseaseorAttack: num [1:253680] 0 0 0 0 0 0 0 0 1 0 ...
 $ PhysActivity         : num [1:253680] 0 1 0 1 1 1 0 1 0 0 ...
 $ Fruits               : num [1:253680] 0 0 1 1 1 1 0 0 1 0 ...
 $ Veggies              : num [1:253680] 1 0 0 1 1 1 0 1 1 1 ...
 $ HvyAlcoholConsump    : num [1:253680] 0 0 0 0 0 0 0 0 0 0 ...
 $ AnyHealthcare        : num [1:253680] 1 0 1 1 1 1 1 1 1 ...
 $ NoDocbcCost          : num [1:253680] 0 1 1 0 0 0 0 0 0 0 ...
 $ GenHlth              : num [1:253680] 5 3 5 2 2 2 3 3 5 2 ...
 $ MentHlth             : num [1:253680] 18 0 30 0 3 0 0 0 30 0 ...
 $ PhysHlth             : num [1:253680] 15 0 30 0 0 2 14 0 30 0 ...
 $ Diffwalk             : num [1:253680] 1 0 1 0 0 0 0 1 1 0 ...
 $ Sex                  : num [1:253680] 0 0 0 0 0 1 0 0 0 1 ...
 $ Age                  : num [1:253680] 9 7 9 11 11 10 9 11 9 8 ...
 $ Education            : num [1:253680] 4 6 4 3 5 6 6 4 5 4 ...
 $ Income               : num [1:253680] 3 1 8 6 4 8 7 4 1 3 ...

```

Датасет до очищення

З метою очищення даних було проведено перетворення деяких змінних, зокрема категоріальних та цензурованих. Значення цих змінних приведено у відповідність до назв у кодовій книжці (codebook). Задля цього було використано скрипт на мові “R”, який зчитує, перетворює та зберігає у файлі “.csv” результат роботи.

Файл очищених даних містить 253 680 записів із 22 змінними, пропущених даних немає.



Rows: 253680 Columns: 22

— Column specification —

Delimiter: ",",

dbl (22): Diabetes_binary, HighBP, HighChol, CholCheck, BMI, Smoker, Stroke, HeartDiseaseorAttack, PhysAc...

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show_col_types = FALSE` to quiet this message.

tibble [253,680 × 22] (S3: tbl_df/tbl/data.frame)

```
$ Diabetes      : logi [1:253680] FALSE FALSE FALSE FALSE FALSE ...
$ HighBP       : logi [1:253680] TRUE FALSE TRUE TRUE TRUE TRUE ...
$ HighChol     : logi [1:253680] TRUE FALSE TRUE FALSE TRUE TRUE ...
$ CholCheck    : logi [1:253680] TRUE FALSE TRUE TRUE TRUE TRUE ...
$ BMI          : num [1:253680] 40 25 28 27 24 25 30 25 30 24 ...
$ Smoker       : logi [1:253680] TRUE TRUE FALSE FALSE FALSE TRUE ...
$ Stroke       : logi [1:253680] FALSE FALSE FALSE FALSE FALSE ...
$ HeartDiseaseorAttack: logi [1:253680] FALSE FALSE FALSE FALSE FALSE ...
$ PhysActivity : logi [1:253680] FALSE TRUE FALSE TRUE TRUE TRUE ...
$ Fruits       : logi [1:253680] FALSE FALSE TRUE TRUE TRUE TRUE ...
$ Veggies      : logi [1:253680] TRUE FALSE FALSE TRUE TRUE TRUE ...
$ HvyAlcoholConsump : logi [1:253680] FALSE FALSE FALSE FALSE FALSE ...
$ AnyHealthcare : logi [1:253680] TRUE FALSE TRUE TRUE TRUE TRUE ...
$ NoDocbcCost  : logi [1:253680] FALSE TRUE TRUE FALSE FALSE ...
$ GenHlth      : Ord.factor w/ 5 levels "poor"<"fair"<...: 1 3 1 4 4 4 3 3 1 4 ...
$ Menthlth     : num [1:253680] 18 0 30 0 3 0 0 0 30 0 ...
$ PhysHlth     : num [1:253680] 15 0 30 0 0 2 14 0 30 0 ...
$ Diffwalk     : logi [1:253680] TRUE FALSE TRUE FALSE FALSE ...
$ Sex          : Factor w/ 2 levels "female","male": 1 1 1 1 1 2 1 1 1 2 ...
$ Age          : Ord.factor w/ 13 levels "18-24"<"25-29"<...: 9 7 9 11 11 10 9 11 9 8 ...
$ Education    : Ord.factor w/ 6 levels "No education"<...: 4 6 4 3 5 6 6 4 5 4 ...
$ Income       : Ord.factor w/ 8 levels "Less than $10,000"<...: 3 1 8 6 4 8 7 4 1 3 ...
```

Датасет після очищення

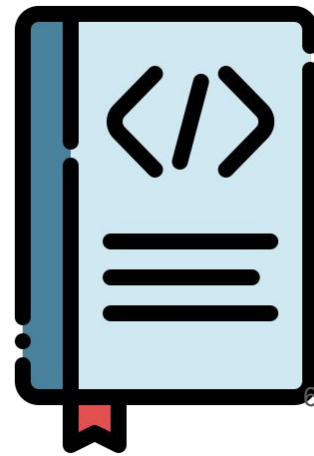
Пояснення щодо особливостей окремих змінних. Окремі витримки з кодової книги (Codebook).

Інтуїтивно зрозумілі змінні:

Diabetes, HighBP, HighChol, BMI, Stroke, HeartDiseaseorAttack, AnyHealthcare, Income, Education, GenHlth

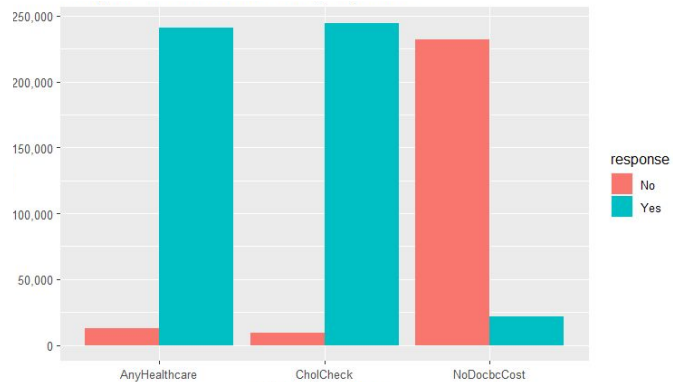
Потребують пояснення:

- CholCheck : 0 = не було жодної перевірки рівня холестерину за останні 5 років, 1 = була хоча б одна перевірка рівня холестерину;
- Smoker : Чи викурили ви принаймні 100 сигарет за все своє життя? [Примітка: 5 пачок = 100 сигарет] 0 = ні, 1 = так;
- PhysActivity : Фізична активність за останні 30 днів - не враховуючи роботу: 0 = ні, 1 = так;
- Fruits : Вживаєте фрукти 1 або більше разів на день: 0 = ні, 1 = так;
- Veggies : Вживаєте овочі 1 або більше разів на день: 0 = ні, 1 = так;
- HvyAlcoholConsump: (дорослі чоловіки ≥ 14 напоїв на тиждень та дорослі жінки ≥ 7 напоїв на тиждень) : 0 = ні, 1 = так;
- NoDocbcCost : Чи був час за останні 12 місяців, коли Вам потрібно було звернутися до лікаря, але Ви не змогли цього зробити через вартість послуг? 0 = ні, 1 = так;
- MentHlth : Шкала днів поганого психічного здоров'я 1-30 днів;
- PhysHlth : Дні фізичної хвороби або травми за останні 30 днів за шкалою від 1 до 30;

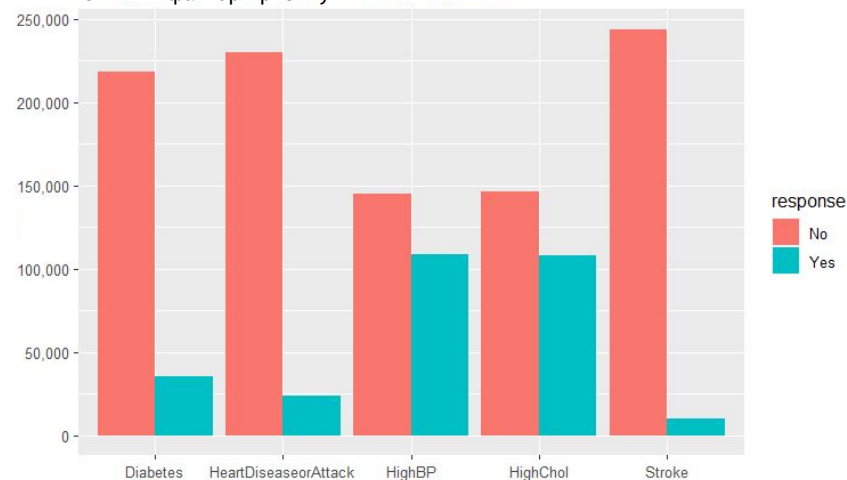


Особливості розподілів окремих змінних Бінарні змінні

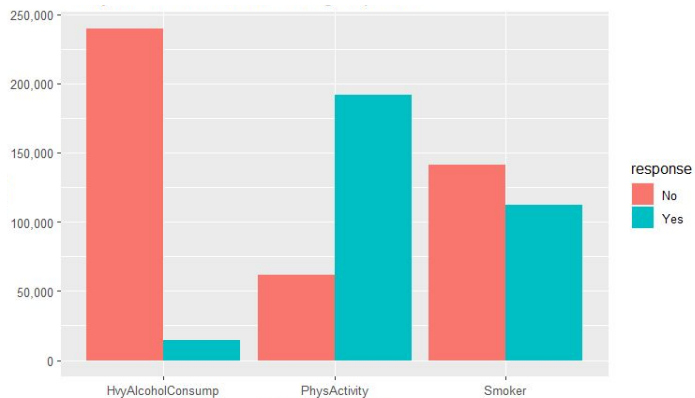
Змінні, що стосуються рівня заробітку та інше



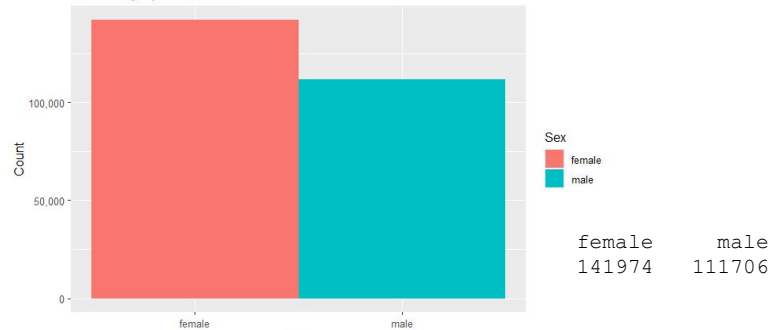
Змінні - фактори ризику



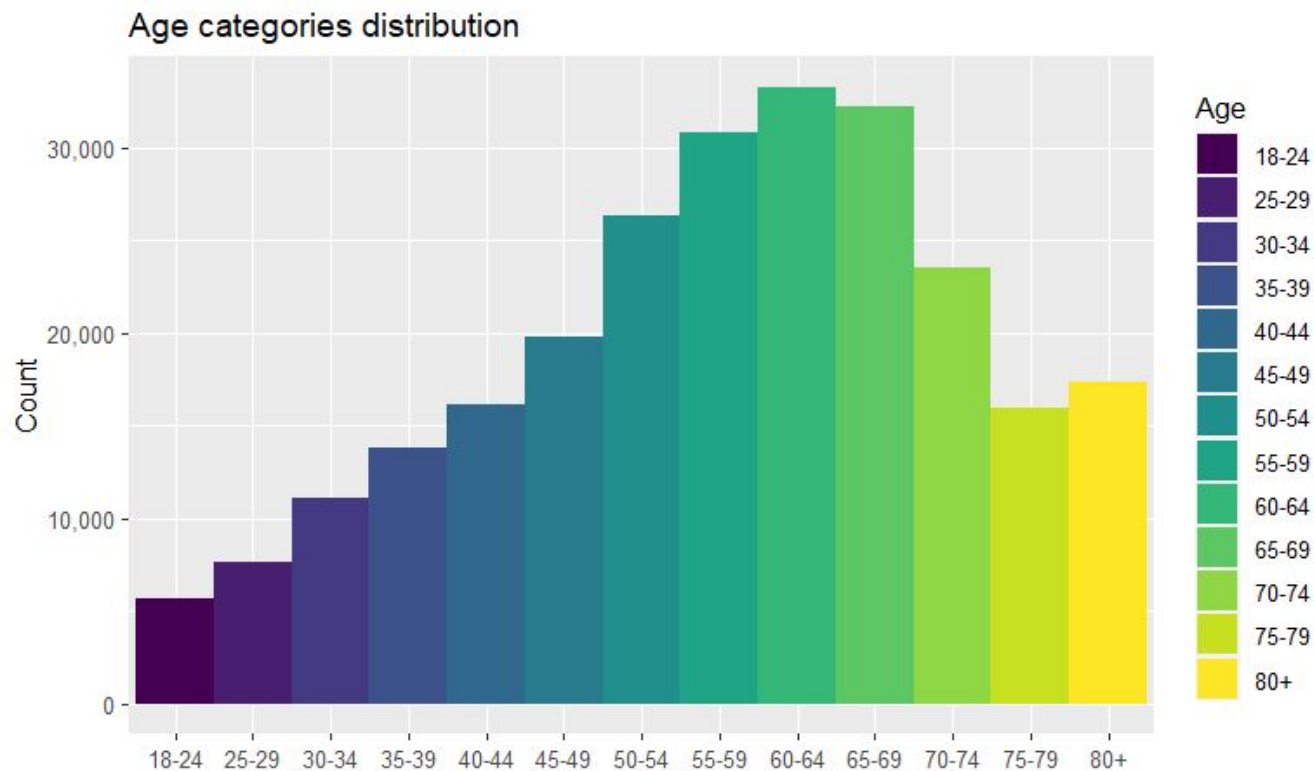
Змінні, що стосуються способу життя



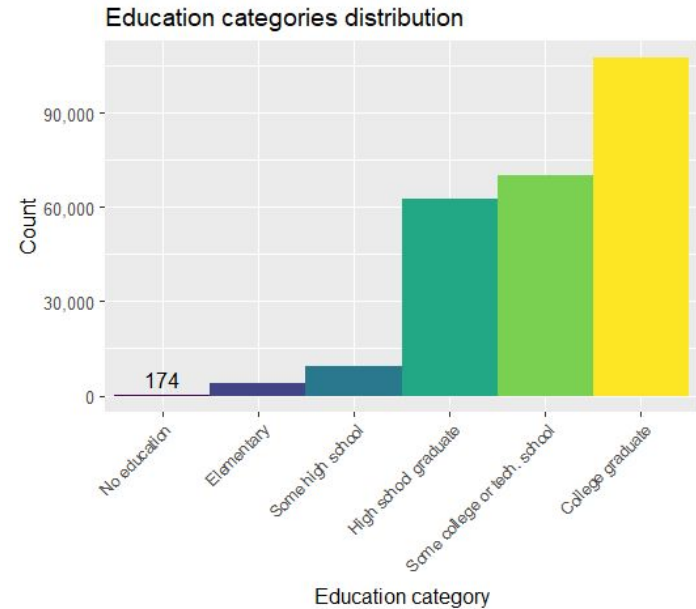
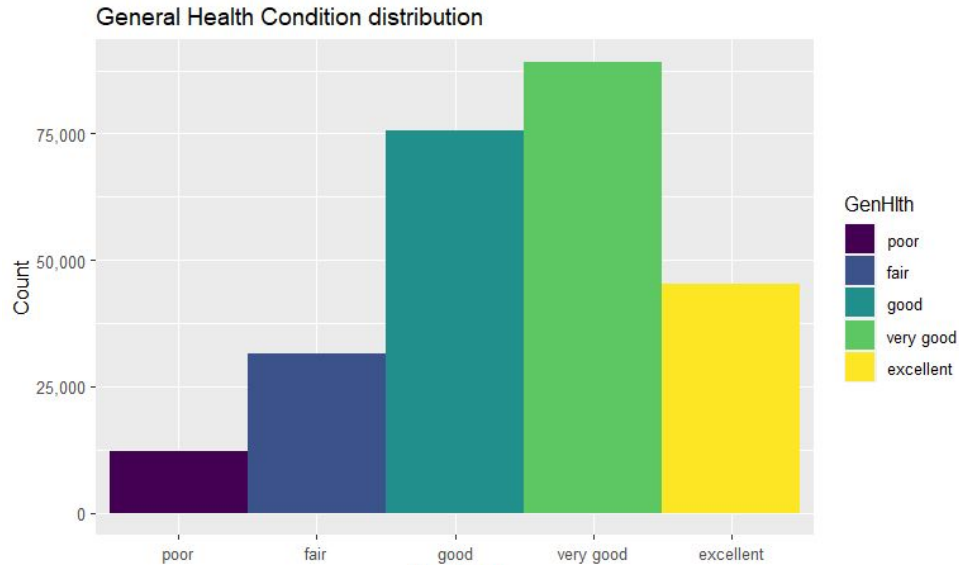
Sex category distribution



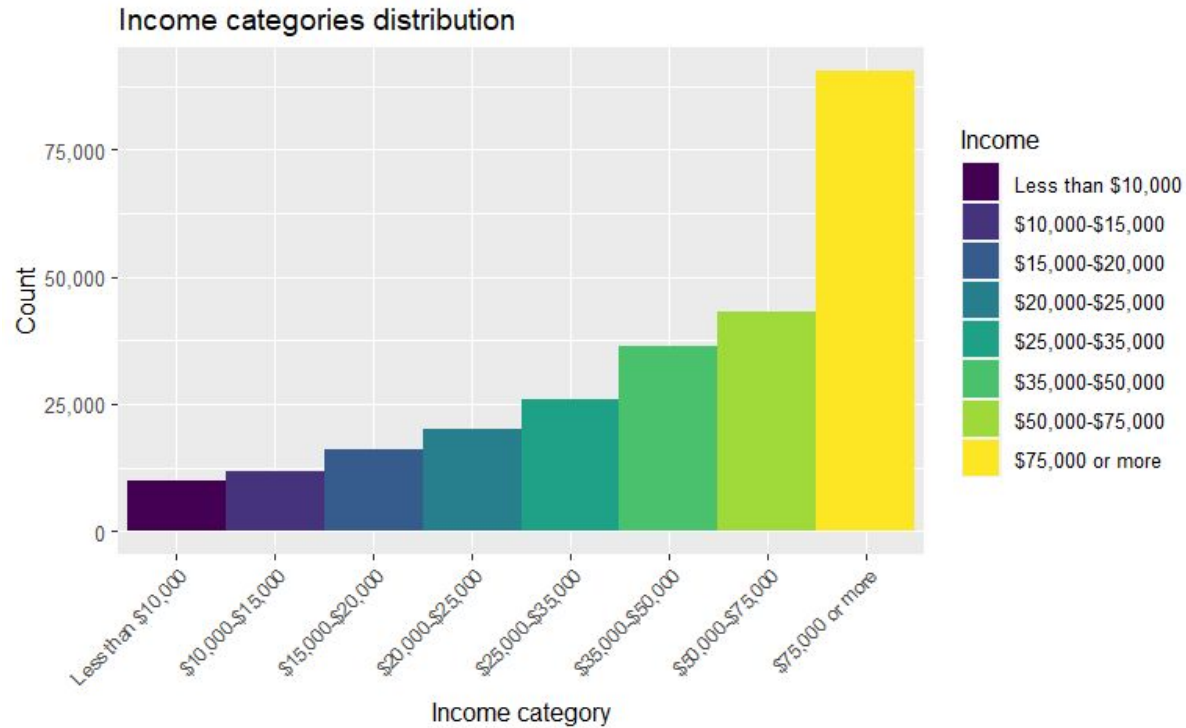
Особливості розподілів окремих змінних Впорядковані категорійні змінні



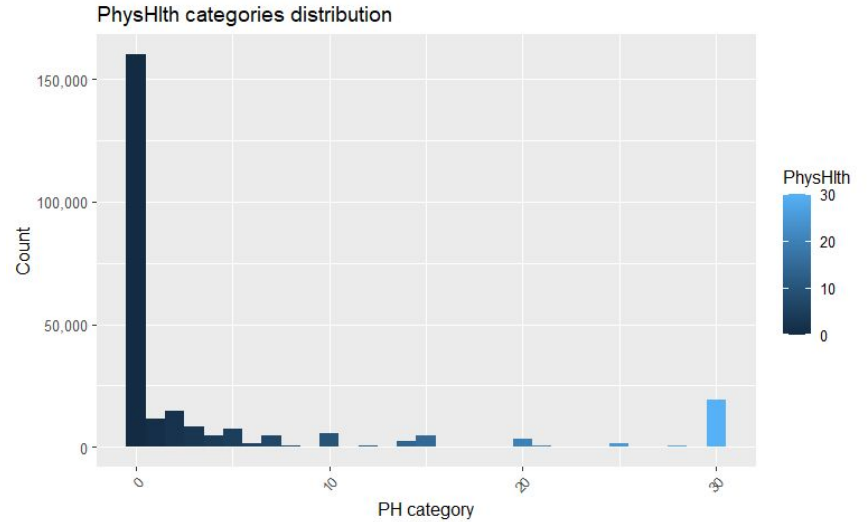
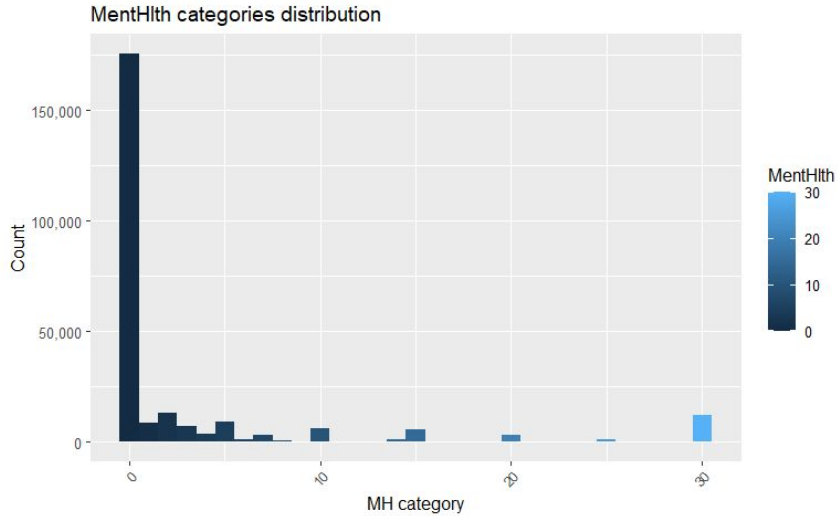
Особливості розподілів окремих змінних Впорядковані категорійні змінні



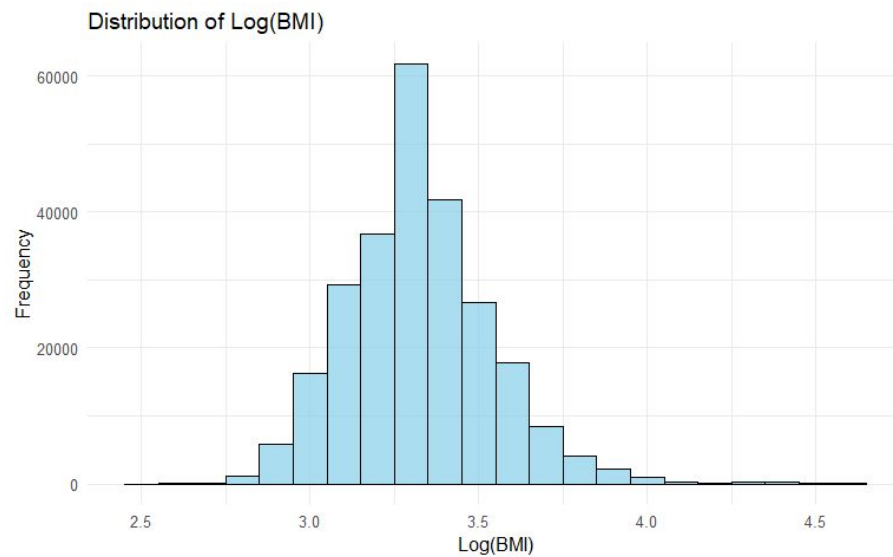
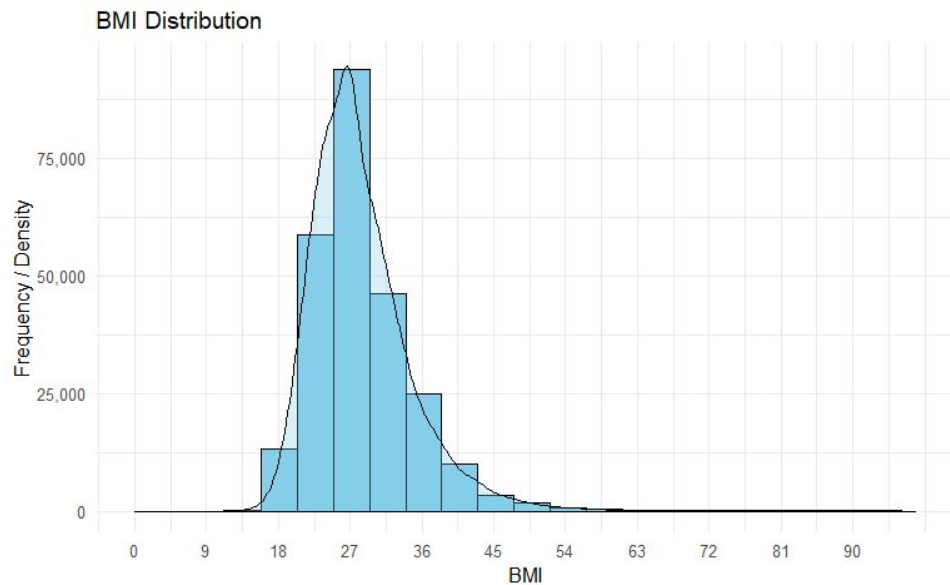
Особливості розподілів окремих змінних Впорядковані категорійні змінні



Особливості розподілів окремих змінних Впорядковані категорійні змінні



Особливості розподілів окремих змінних. Неперервні змінні.

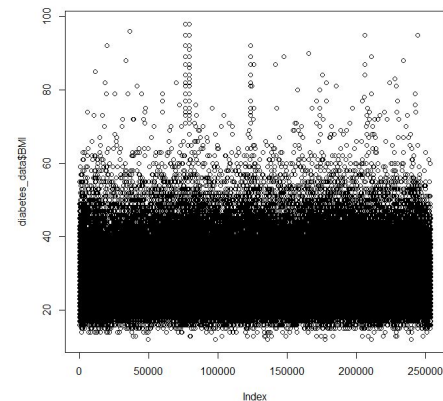
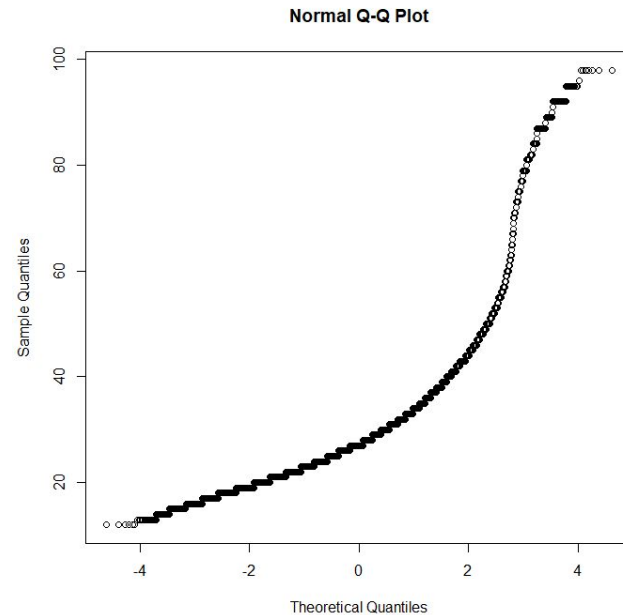


Дескриптивні статистики окремих змінних. IMT

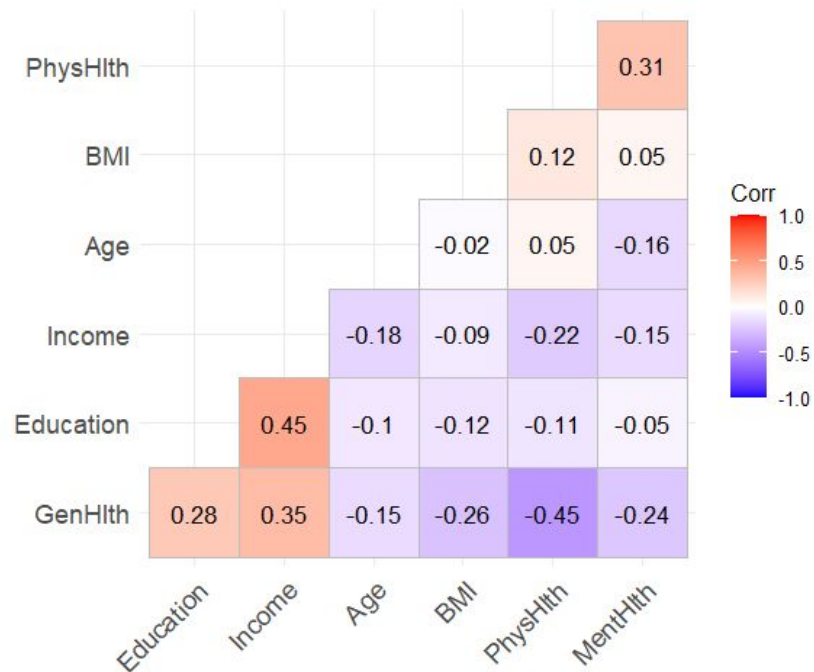
- Середнє: 28.38236
- Середньоквадратичне відхилення: 6.608694
- Підсумкові характеристики:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12.00	24.00	27.00	28.38	31.00	98.00

- За правилом трьох сигм виявлено 2963 викидів, інтервал - [12; 48]
- Фільтр Гампеля показує 11506 викидів, залишаючи інтервал [14; 40]
- На qqplot не виявлено істотних викидів (див. додаток А звіту)
- На діаграмі розкиду - так само, за винятком кількох значень 98 і 96



Візуалізація коефіцієнтів кореляції Спірмена для пар впорядкованих факторів/неперервних змінних.



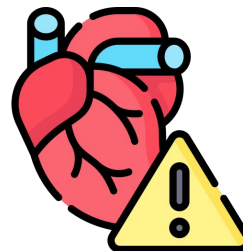
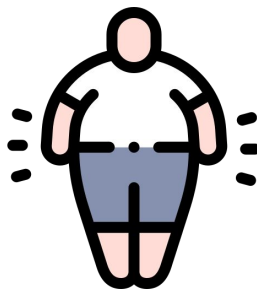
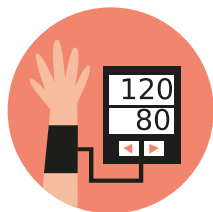
Питання 1

Залежність наявності діабету від факторів ризику (підвищений рівень холестерину, наявність ожиріння, проблеми з серцево-судинною системою, вік)

Опис питання

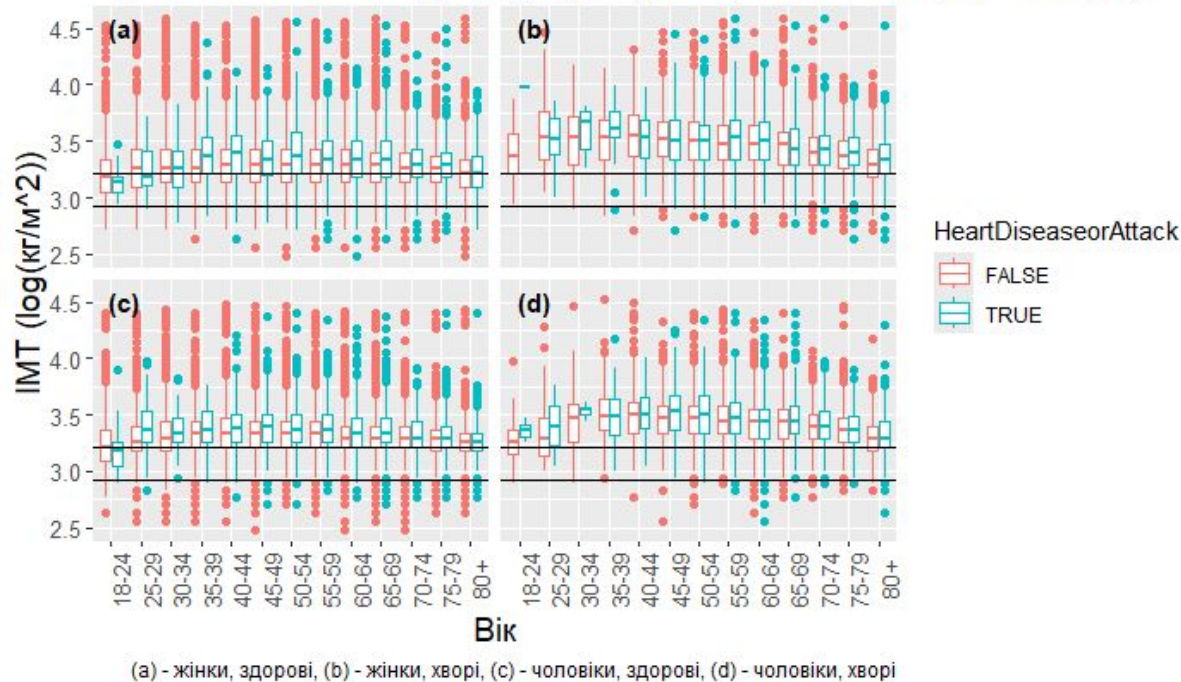
Суть питання полягає в тому, щоб з'ясувати, чи може діабет бути зумовленим несприятливими факторами здоров'я, а саме - завищеним або заниженим ІМТ, високим рівнем холестерину, артеріальним тиском, пережитим інсультом або віком.

Питання цікаве, тому що відповідь на нього може уточнити міру впливу того чи іншого фактора, або виявити, що певний фактор не впливає на наявність діабету.



Графік 1

Залежність наявності діабету від віку, статі, ІМТ та захворювань серця



- На графіку видно, що для хворих на діабет загалом спостерігаються більші значення основних підсумкових характеристик
- Медіанне значення ІМТ часто перевищує здорові межі
- ІМТ у хворих на діабет із захворюваннями серця та без них мало відрізняються
- Люди старшого віку більш схильні до наявності захворювань серця та діабету, ніж молоді
- Чоловіки більш схильні до хвороб, ніж жінки

- Таблиця співвідношень кількості людей за наявністю хвороб

, , Diabetes = FALSE, Sex = female

		Age													
HeartDiseaseorAttack		18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	
FALSE		2689	3879	5807	7271	8436	9807	12727	14572	14491	13539	9603	6653	7242	
TRUE		11	23	64	88	124	219	436	642	812	954	1003	931	1540	

, , Diabetes = TRUE, Sex = female

	Age													
HeartDiseaseorAttack	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	
FALSE	44	83	183	348	540	784	1419	1914	2445	2664	1992	1415	1222	
TRUE	1	6	8	18	36	118	223	341	523	586	561	419	518	

, , Diabetes = FALSE, Sex = male

	Age													
HeartDiseaseorAttack	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	
FALSE	2907	3533	4888	5771	6402	7792	9551	10494	10886	9438	6278	3821	3842	
TRUE	15	23	50	67	144	259	512	861	1322	1705	1508	1172	1530	

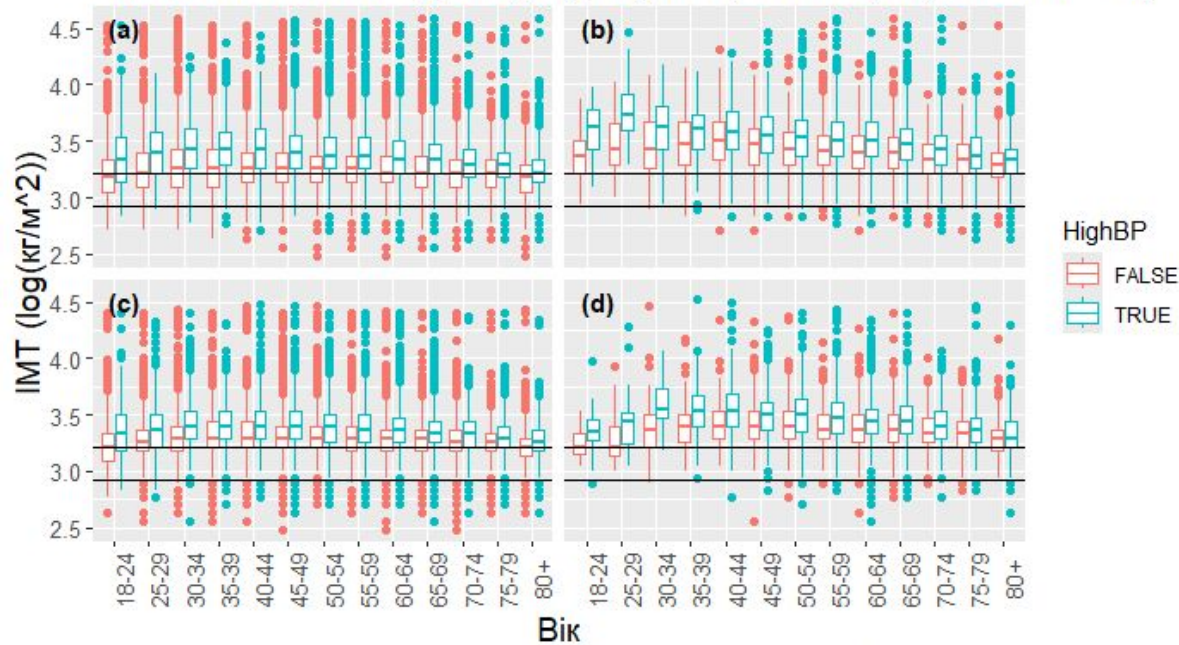
, , Diabetes = TRUE, Sex = male

	Age													
HeartDiseaseorAttack	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	
FALSE	31	49	119	240	428	724	1192	1599	2064	2360	1713	998	898	
TRUE	2	2	4	20	47	116	254	409	701	948	875	571	571	

- Більшість людей із захворюваннями серця не має діабету

Графік 2

Залежність наявності діабету від віку, статі, IMT та артеріального тиску



(a) - жінки, здорові, (b) - жінки, хворі, (c) - чоловіки, здорові, (d) - чоловіки, хворі

- Спостерігається істотна різниця між IMT у тих, хто має високий артеріальний тиск, і тих, хто його не має
- На цю різницю майже не впливає наявність діабету
- У молодших вікових груп видно значний відрив між людьми з артеріальним тиском та без нього

- Таблиця співвідношень кількості людей за наявністю хвороб

, , Diabetes = FALSE, Sex = female

	Age												
HighBP	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
FALSE	2556	3632	5274	6438	7118	7837	9526	10036	9129	7497	4743	2905	3124
TRUE	144	270	597	921	1442	2189	3637	5178	6174	6996	5863	4679	5658

, , Diabetes = TRUE, Sex = female

	Age												
HighBP	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
FALSE	35	63	117	189	251	331	529	590	664	637	447	320	333
TRUE	10	26	74	177	325	571	1113	1665	2304	2613	2106	1514	1407

, , Diabetes = FALSE, Sex = male

	Age												
HighBP	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
FALSE	2594	2994	4019	4522	4850	5615	6373	6446	6313	5128	3218	1975	2247
TRUE	328	562	919	1316	1696	2436	3690	4909	5895	6015	4568	3018	3125

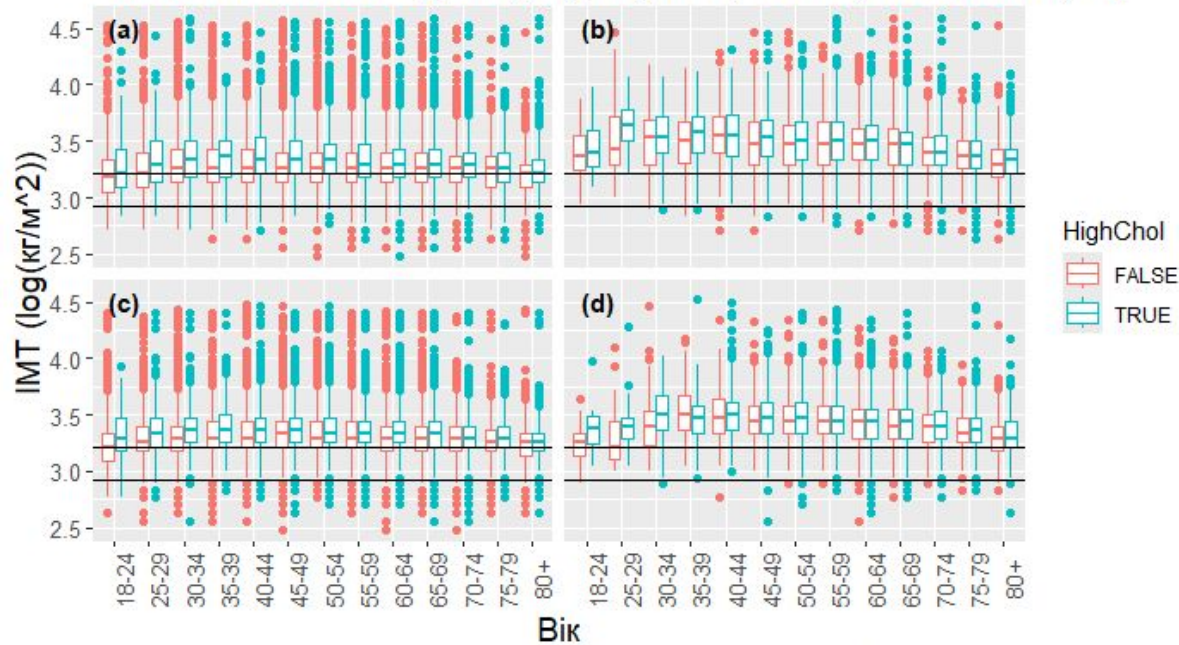
, , Diabetes = TRUE, Sex = male

	Age												
HighBP	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
FALSE	23	28	77	105	197	301	427	511	620	680	548	348	371
TRUE	10	23	46	155	278	539	1019	1497	2145	2628	2040	1221	1098

- Є разюча різниця у співвідношенні людей із тиском та без нього при діабеті та за його відсутності
- Навіть при тиску більшість не мають діабету

Графік 3

Залежність наявності діабету від віку, статі, ІМТ та рівня холестерину



(a) - жінки, здорові, (b) - жінки, хворі, (c) - чоловіки, здорові, (d) - чоловіки, хворі

- Ті, хто не має діабету, але має високий рівень холестерину, має також і вищий ІМТ порівняно з тими, у кого він не є високим
- Коробкові графіки у хворих на діабет приблизно однакові як при високому рівню холестерину, так і без нього

- Таблиця співвідношень кількості людей за наявністю хвороб

, , Diabetes = FALSE, Sex = female

	Age												
HighChol	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
FALSE	2456	3445	5077	6146	6837	7431	8947	9212	8327	7171	4900	3566	4297
TRUE	244	457	794	1213	1723	2595	4216	6002	6976	7322	5706	4018	4485

, , Diabetes = TRUE, Sex = female

	Age												
HighChol	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
FALSE	30	59	109	186	240	347	548	683	850	908	735	577	618
TRUE	15	30	82	180	336	555	1094	1572	2118	2342	1818	1257	1122

, , Diabetes = FALSE, Sex = male

	Age												
HighChol	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
FALSE	2687	3114	4039	4438	4598	5160	6060	6301	6287	5257	3493	2394	2789
TRUE	235	442	899	1400	1948	2891	4003	5054	5921	5886	4293	2599	2583

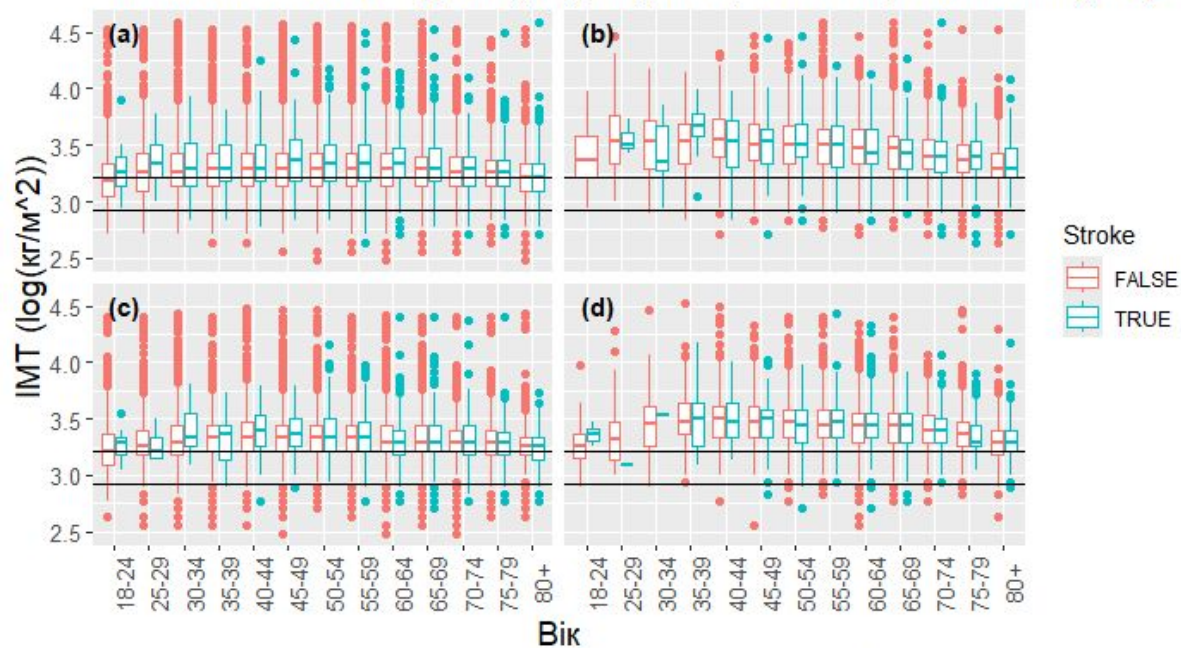
, , Diabetes = TRUE, Sex = male

	Age												
HighChol	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
FALSE	25	27	70	106	197	316	482	646	891	1015	853	563	579
TRUE	8	24	53	154	278	524	964	1362	1874	2293	1735	1006	890

- Ті, хто мають діабет, найімовірніше мають і стороннє захворювання, але не навпаки

Графік 4

Залежність наявності діабету від віку, статі, IMT та пережитого інсульту



(a) - жінки, здорові, (b) - жінки, хворі, (c) - чоловіки, здорові, (d) - чоловіки, хворі

- Пережитий інсульт не є гарантією діабету

- Таблиця співвідношень кількості людей за наявністю хвороб

, , Diabetes = FALSE, Sex = female

	Age													
Stroke	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	
FALSE	2691	3883	5820	7290	8436	9870	12865	14785	14851	13971	10046	7083	8017	
TRUE	9	19	51	69	124	156	298	429	452	522	560	501	765	

, , Diabetes = TRUE, Sex = female

	Age													
Stroke	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	
FALSE	45	86	185	353	546	844	1518	2041	2696	2959	2278	1622	1507	
TRUE	0	3	6	13	30	58	124	214	272	291	275	212	233	

, , Diabetes = FALSE, Sex = male

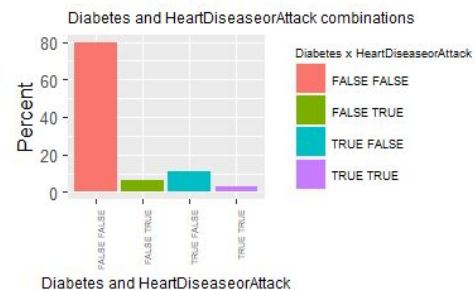
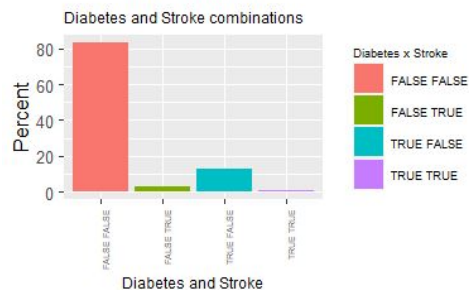
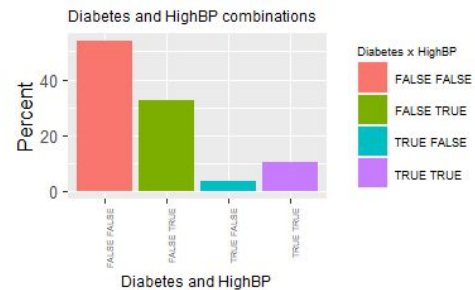
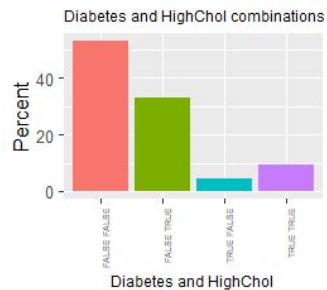
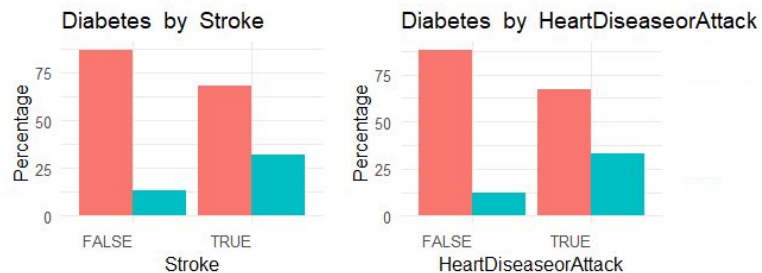
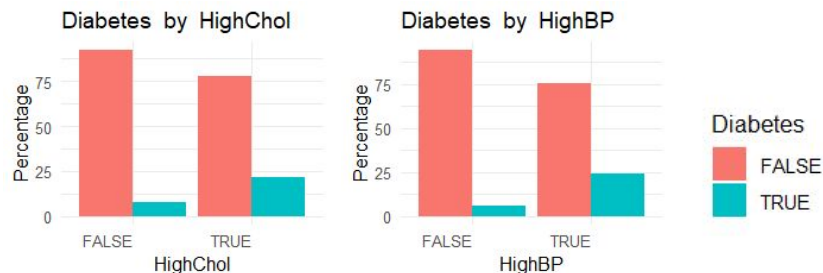
	Age													
Stroke	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	
FALSE	2912	3550	4915	5792	6480	7940	9859	11069	11779	10657	7318	4593	4838	
TRUE	10	6	23	46	66	111	204	286	429	486	468	400	534	

, , Diabetes = TRUE, Sex = male

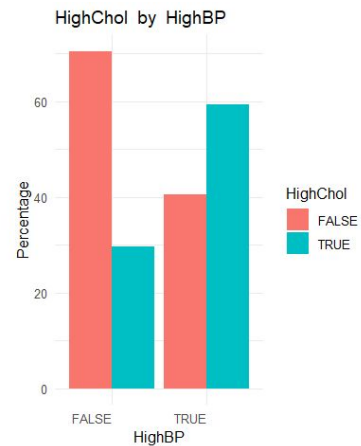
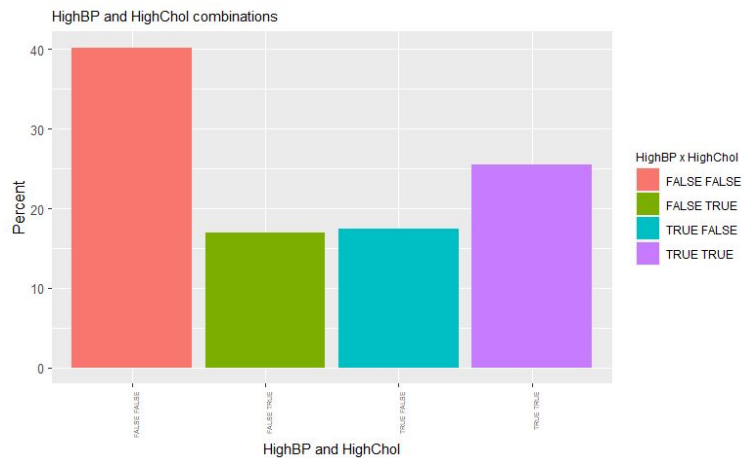
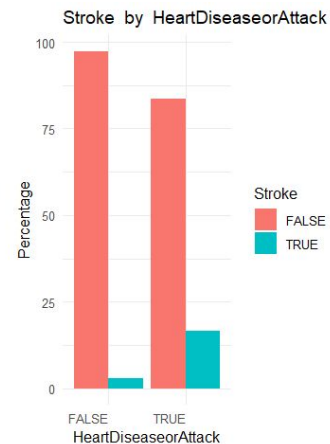
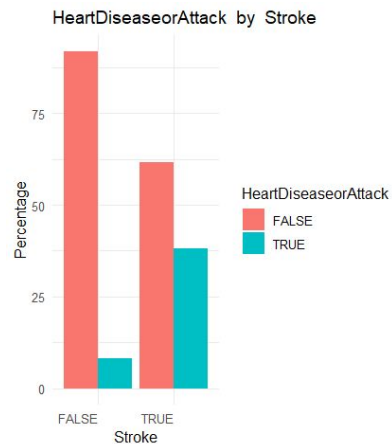
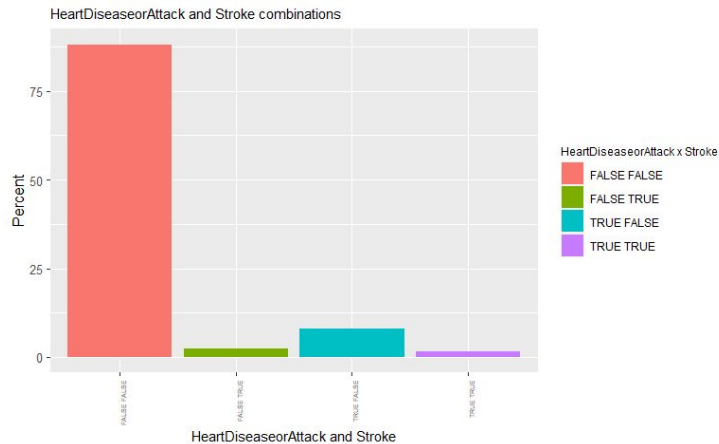
	Age													
Stroke	18-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+	
FALSE	31	50	120	251	466	797	1350	1852	2504	2987	2334	1391	1265	
TRUE	2	1	3	9	9	43	96	156	261	321	254	178	204	

- При діабеті більшість не переживали інсульт, хоча він дещо збільшує ймовірність діабету

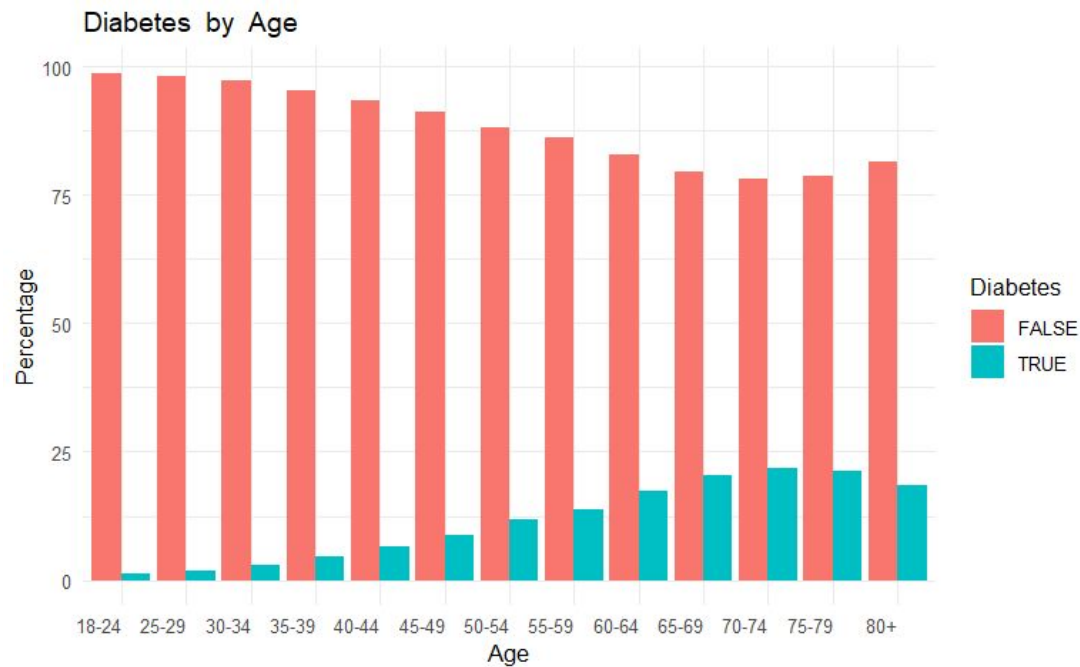
Залежність наявності діабету від бінарних факторів ризику окремо.



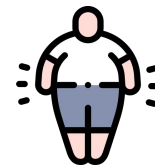
Візуалізації для змінних **Stroke** та **HeartDiseaseorAttack**, **HighBP** та **HighChol** є дещо схожими, тому є сенс перевірити залежність між цими двома змінними окремо.



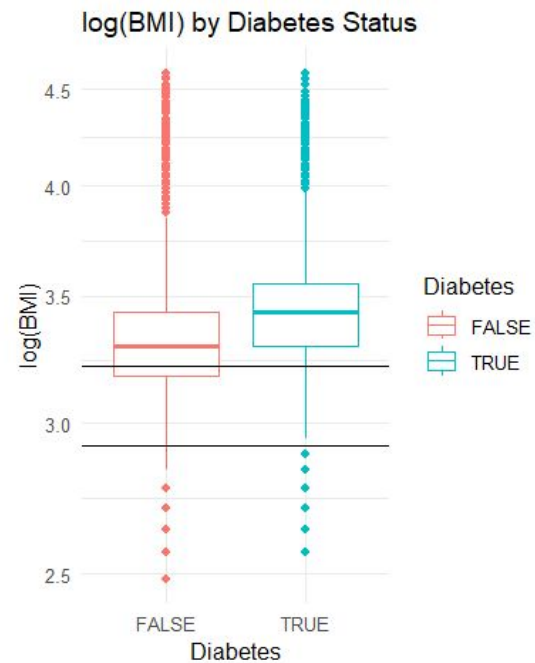
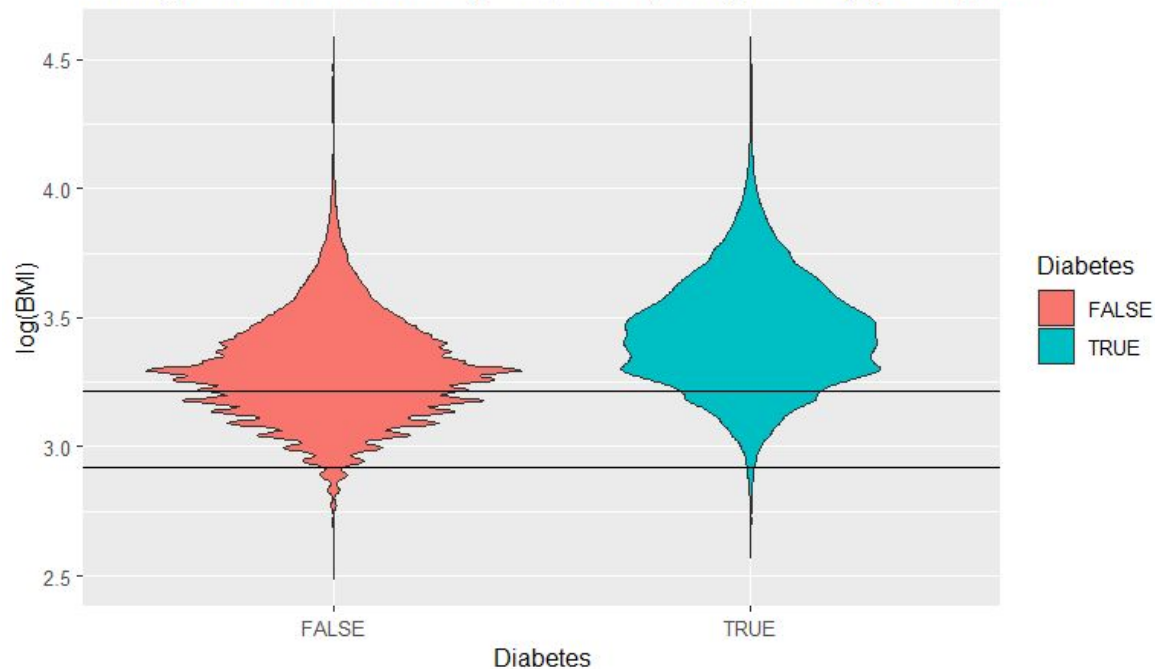
Залежність наявності діабету від вікової категорії



Залежність наявності діабету від ІМТ



Розподіл ІМТ в залежності від того, чи хвора людина на цукровий діабет



Попередні підсумки

- вік (по мірі збільшення) та надмірність ваги є основними факторами наявності діабету
- у жодному випадку побічне захворювання не може однозначно свідчити про діабет
- деякі побічні захворювання (фактори) є тісно пов'язаними одне з одними, тому розглядаючи ізольований вплив цих факторів на наявність у людини діабету (для прогнозування наявності у людини діабету) можна дійти до хибних висновків
- побічні захворювання дещо збільшують ймовірність діабету, проте більшість все одно його не мають
- хворі на діабет найчастіше мають високий рівень тиску та холестерину, особливо, - люди з віком 40+, водночас ці захворювання більш поширені серед людей без діабету
- схоже, що чоловіки більш схильні до діабету, ніж жінки



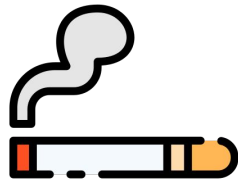
Питання 2

Залежність наявності діабету від соціального статусу (раціон харчування, паління, зарплата, освіта, страховка)

Опис питання

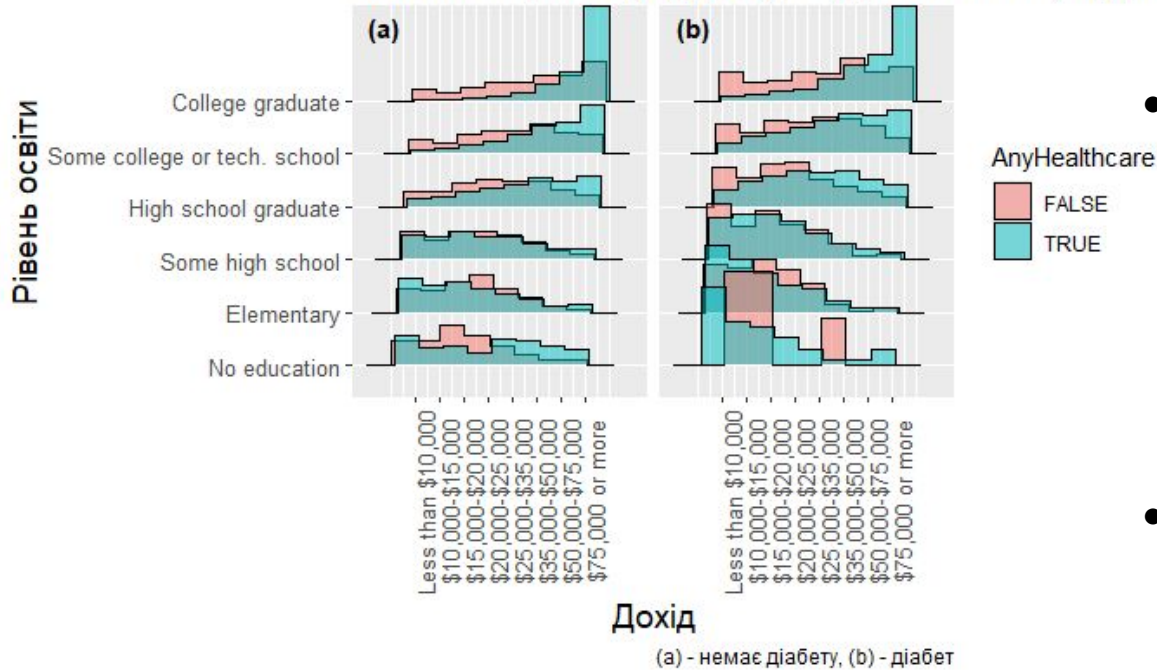
Суть питання полягає в тому, щоб з'ясувати вплив соціально-економічних факторів / способу життя на наявність діабету: харчування, вживання алкоголю, паління, освіти, зарплати і страховки.

Питання цікаве, бо відповідь може пояснити, чи є вплив факторів якості життя на наявність діабету і його характер.



Графік 1

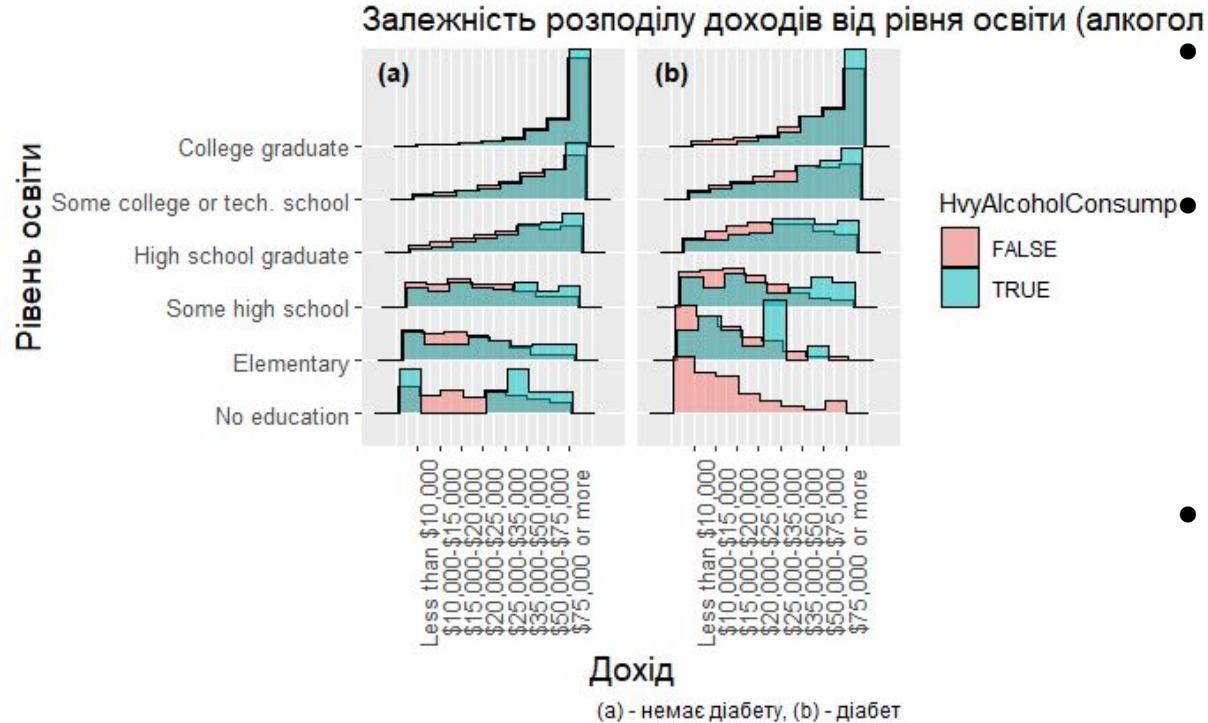
Залежність розподілу доходів від рівня освіти (мед. об



(a) - немає діабету, (b) - діабет

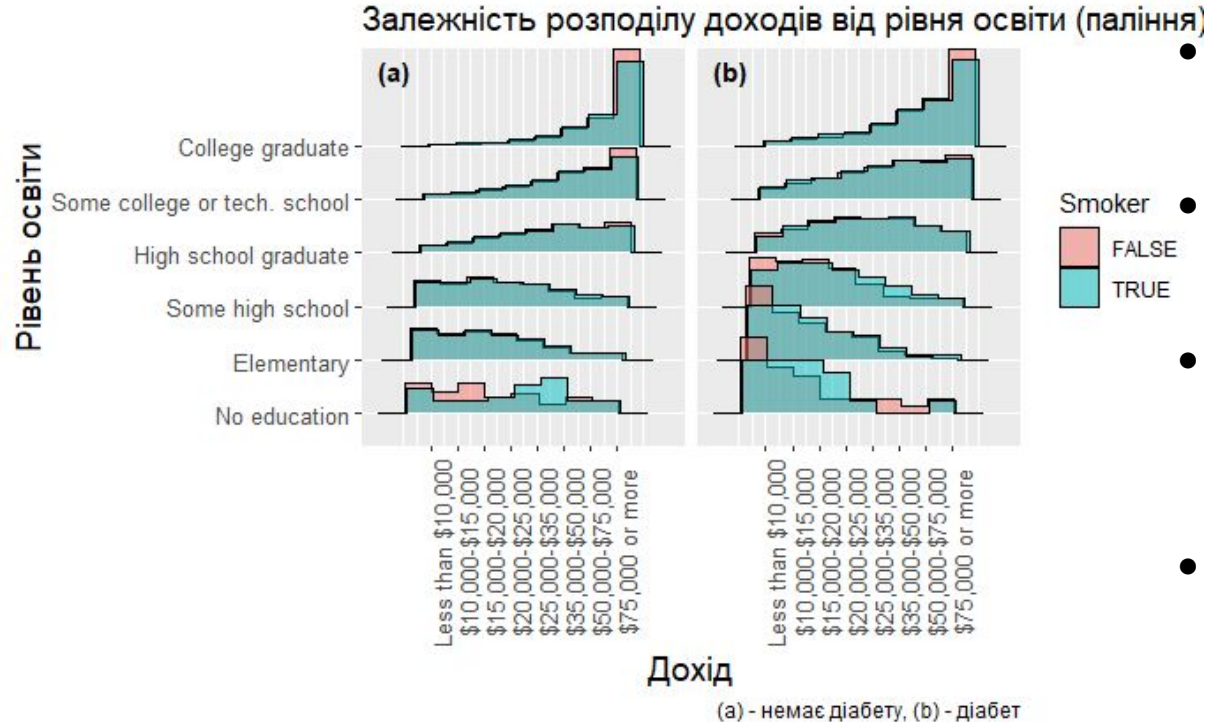
- Видно велику кількість бідних людей без освіти, що мають діабет, тоді як така сама група без діабету майже рівномірно розподілена по доходах
- Видно, що у людей з технічною і вищою освітою розподіл за доходами наближається до рівномірного, якщо вони не користуються медичними послугами, тоді як менш освічені по мірі зменшення доходів схильні не користуватися ними
- Для людей із вищою освітою, діабет може бути пов'язаним із низькими доходами та відсутністю медичного обслуговування, проте в більш освічених не спостерігається сильний вплив цих факторів

Графік 2



- Розподіли подібні як у випадку відсутності діабету, так і за його наявності
- Для людей з діабетом розподіли при відсутності алкоголізму та наявності базової освіти виходять на плато, досягаючи рівня доходів \$20000-25000, тоді як в людей без діабету вони схильні рости
- Серед неосвічених немає таких, що вживають алкоголь та хворіють на діабет. Водночас є певна кількість хворих без алкоголізму

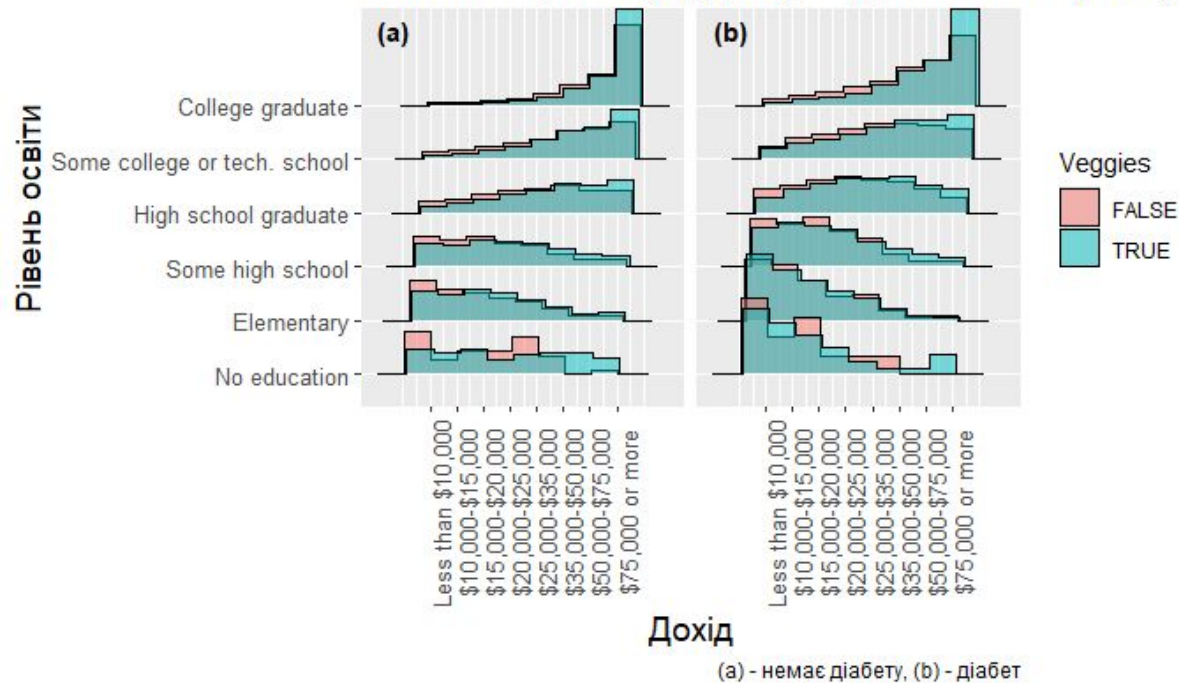
Графік 3



- Розподіли, за винятком людей без освіти, майже ідентичні за критерієм паління
- У декількох груп із низькими доходами спостерігається ріст частки тих, хто не палить
- Серед неосвічених бачимо збільшення частки тих, хто палить і не має діабету при доходах \$25000-50000
- Розподіли прагнуть до вирівнювання при досягненні певного рівня доходів

Графік 4

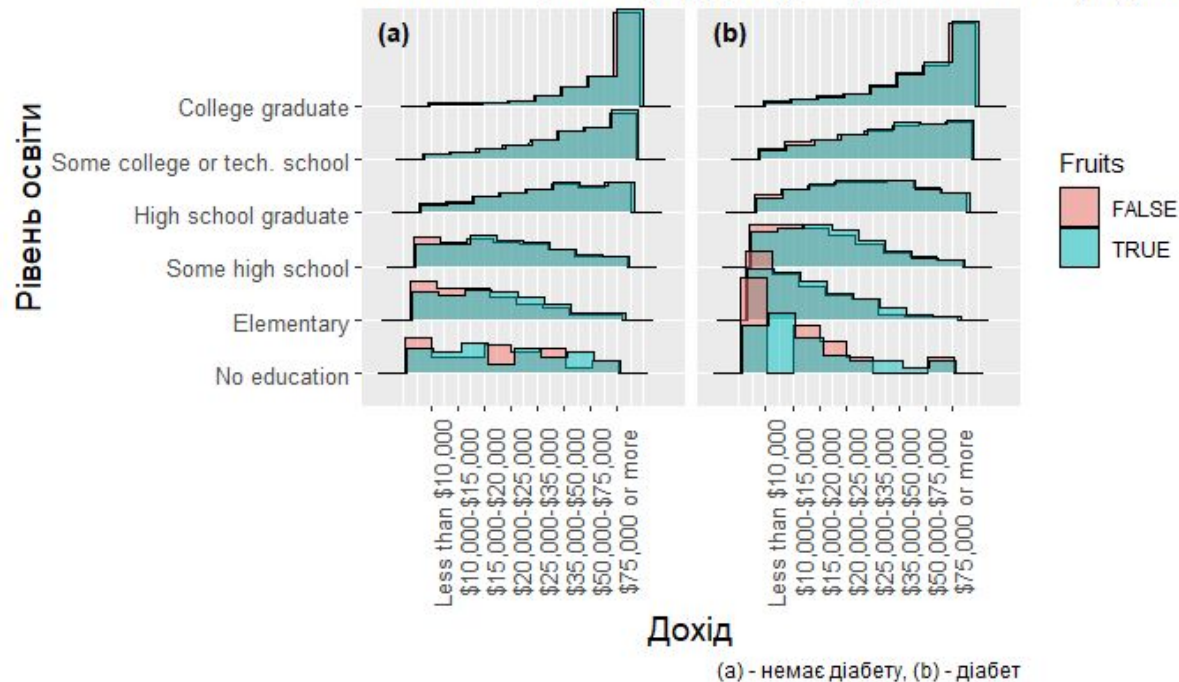
Залежність розподілу доходів від рівня освіти (овочі)



- Можна помітити подібність розподілів за фактором вживання овочів і зменшення приросту при діабеті незалежно від харчування
- Більше багатих схильні вживати овочі, тоді як бідні частіше не вживають їх

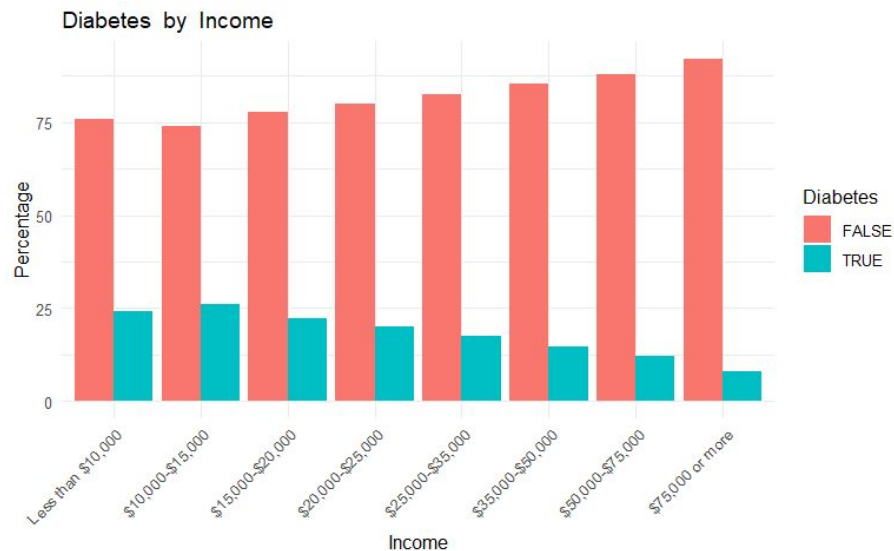
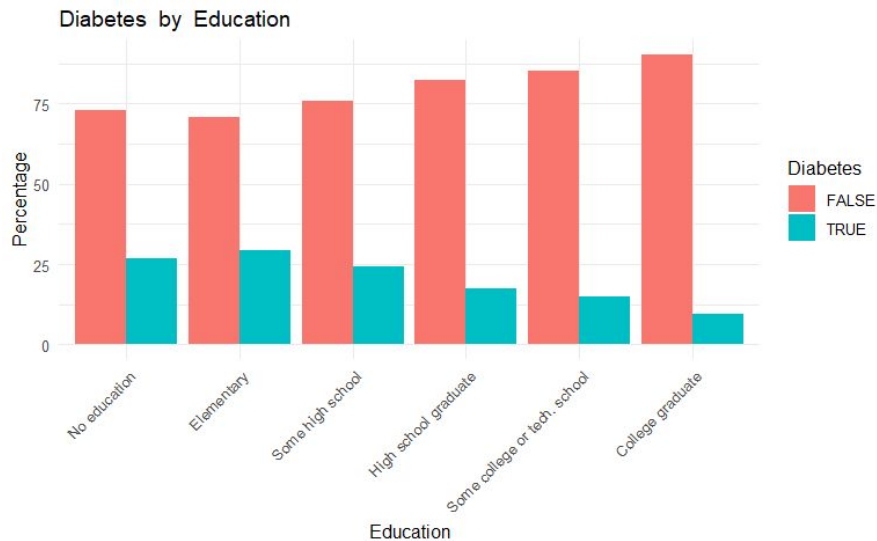
Графік 5

Залежність розподілу доходів від рівня освіти (фрукти)

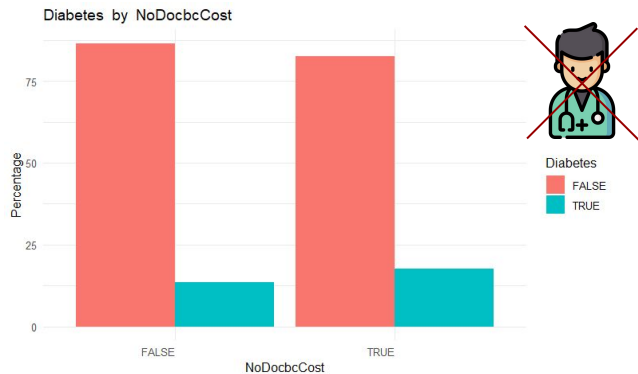
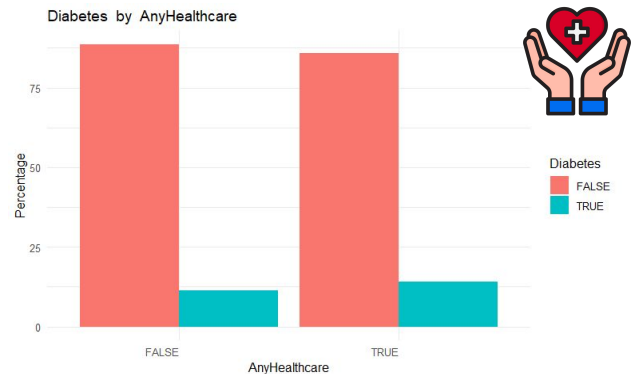


- Розподіли при вживанні фруктів у обох графіках майже не відрізняються від розподілів без їх вживання для тих, хто має базову освіту або вище
- Для груп з нижчим рівнем освіти спостерігається сильний шум в розподілах

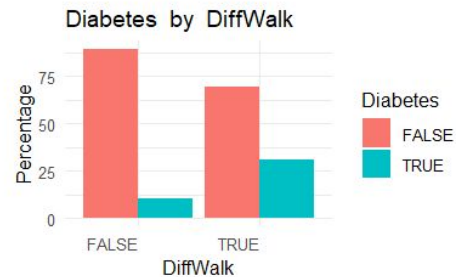
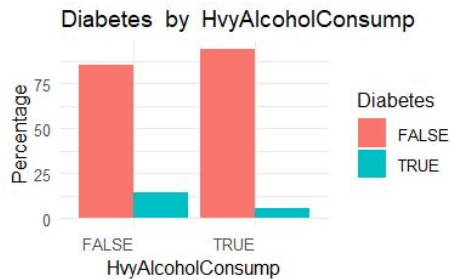
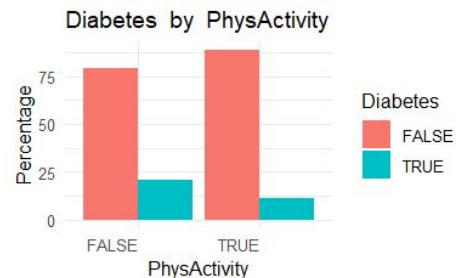
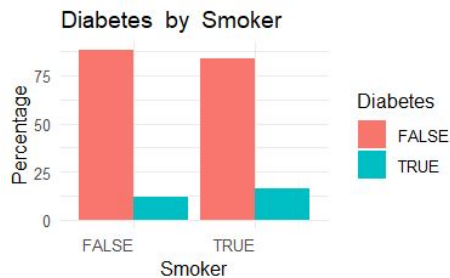
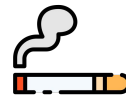
Залежність наявності діабету від рівня освіти, заробітної плати



Залежність наявності діабету від наявності доступу до медичних послуг



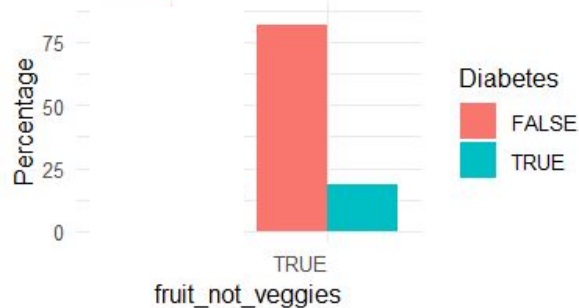
Залежність наявності діабету від способу життя





Залежність наявності діабету від
особливостей харчування.

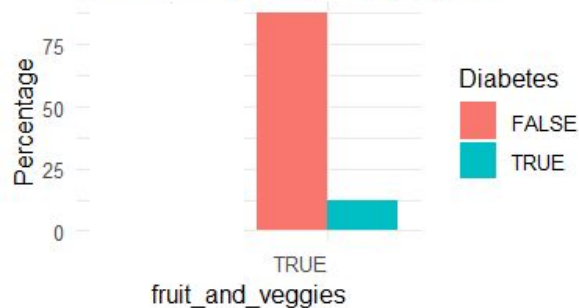
Diabetes by fruit_not_veggies



Diabetes by veggies_not_fruit



Diabetes by fruit_and_veggies



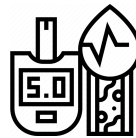
Diabetes by fruit_or_veggies



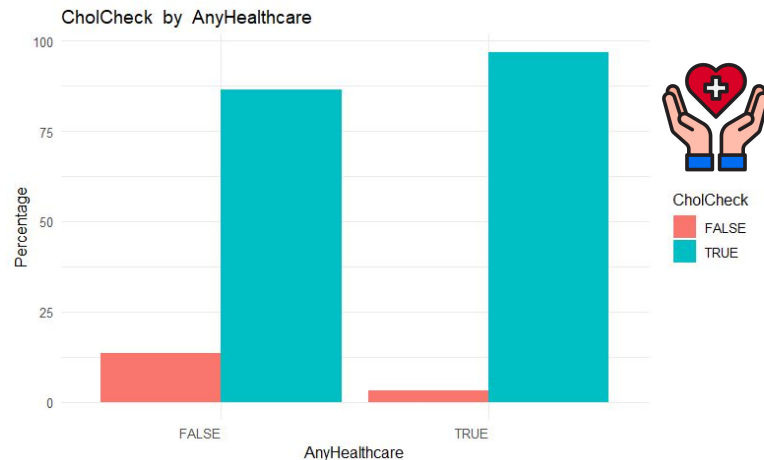
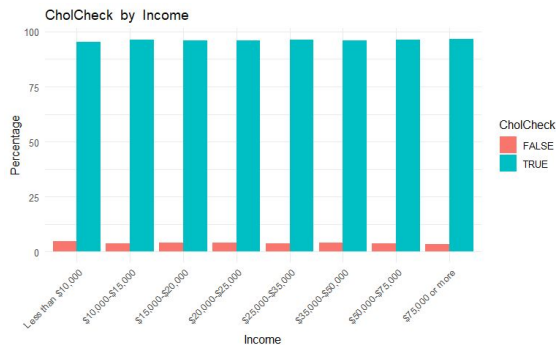
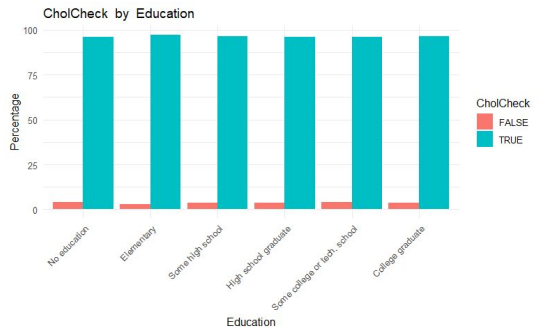
Попередні підсумки

- спостерігаються менші частки хворих на діабет у багатших людей та людей з вищим рівнем освіти
- зі збільшенням доходів деякі із негативних факторів (алкоголь, паління, відсутність в раціоні овочів та фруктів) можуть сприяти діабету
- схоже, що люди без освіти та з низькими доходами частіше мають діабет, але може бути вплив інших факторів (напр. неможливість купити продукти), може даватися в знаки мала кількість таких людей

Питання 3



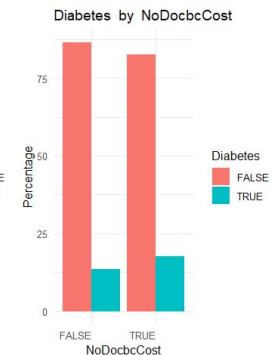
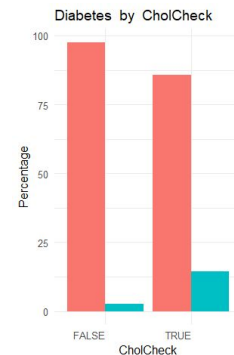
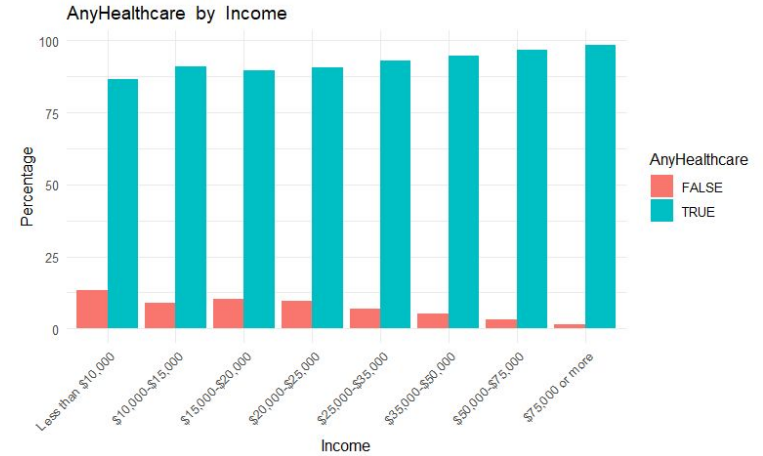
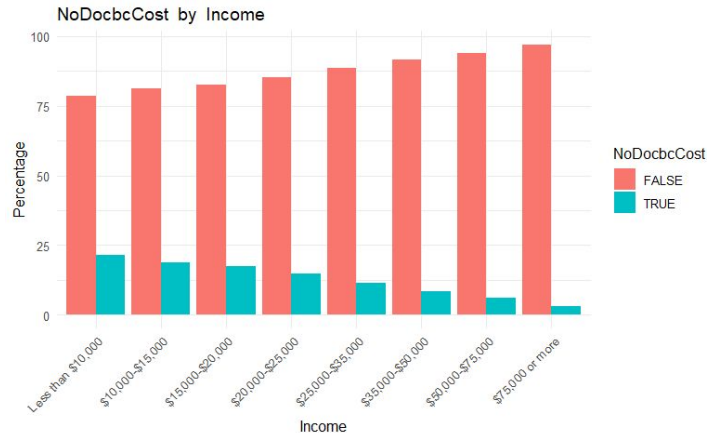
Залежність частоти обстежень рівня холестерину від соціального статусу (рівня освіти, зарплати та медичного страхування)



В результаті аналізу виникло 3 додаткових питання.

В результаті аналізу виникло 3 додаткових питання:

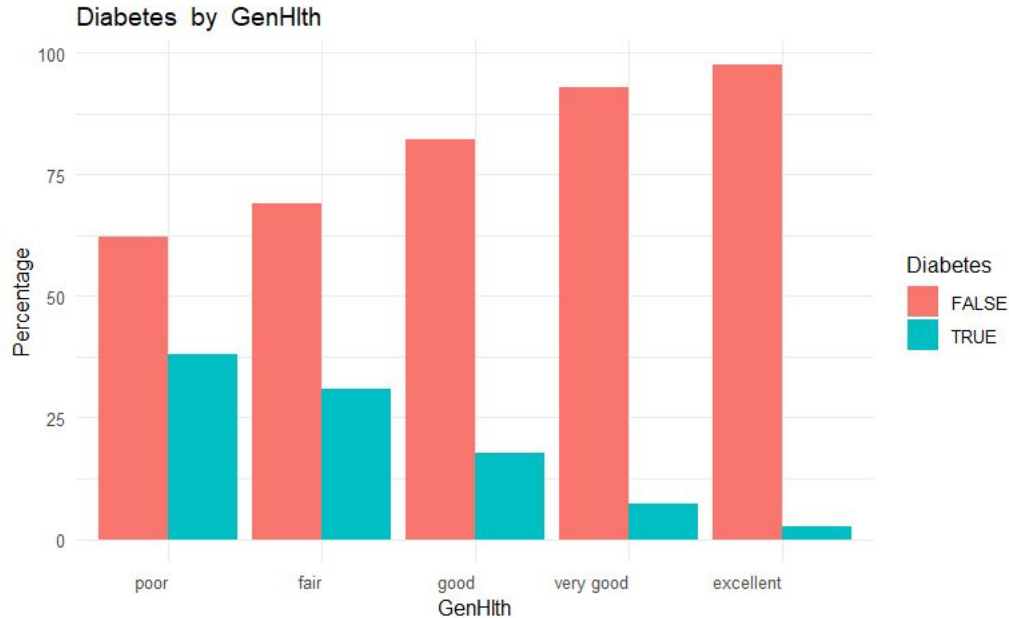
1. Чи є залежність наявності страховки від заробітної плати?
2. Чи є залежність змінної NoDocbcCost від заробітної плати?
3. Залежність наявності діабету від змінних **CholCheck**, **NoDocbcCost**



Питання 4

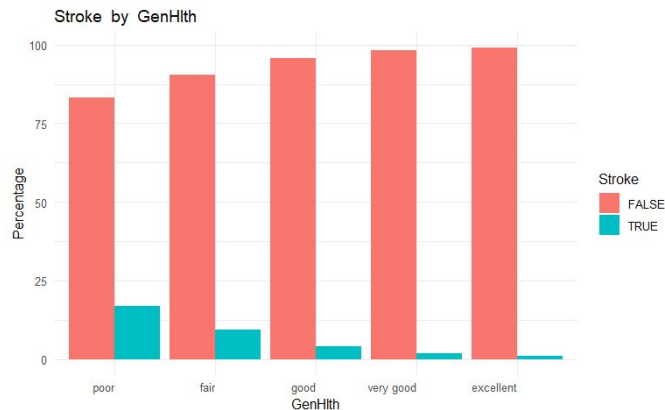
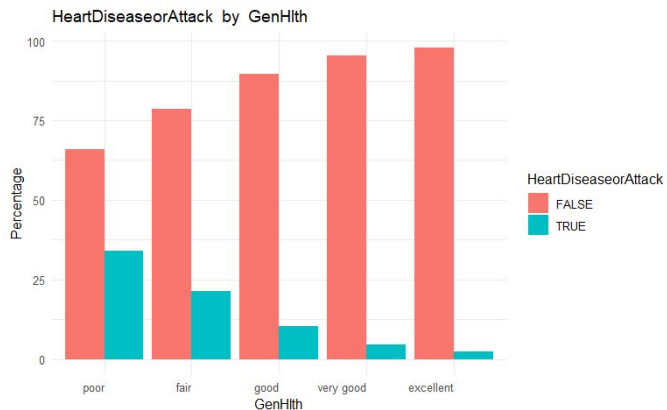
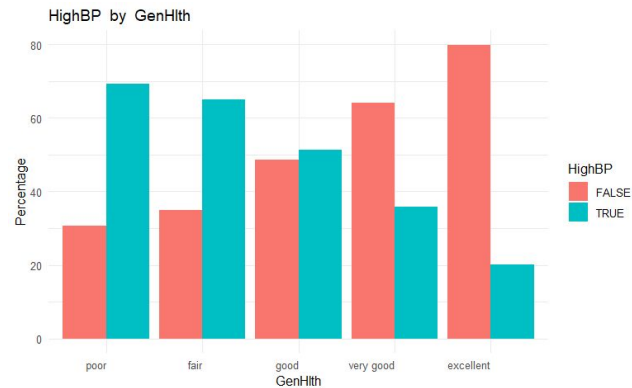
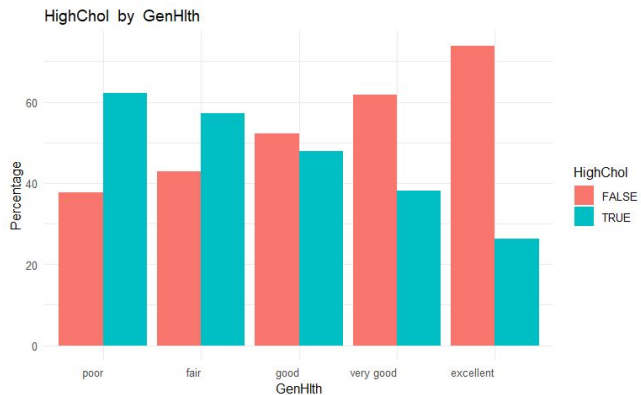
Залежність наявності діабету від суб'єктивної оцінки стану здоров'я.

Порівняння із залежністю наявності діабету від об'єктивних факторів: імт, тиску, чи є хвороби серця і того, чи важко людині ходити (чи завищують / занижують люди оцінку власного здоров'я).

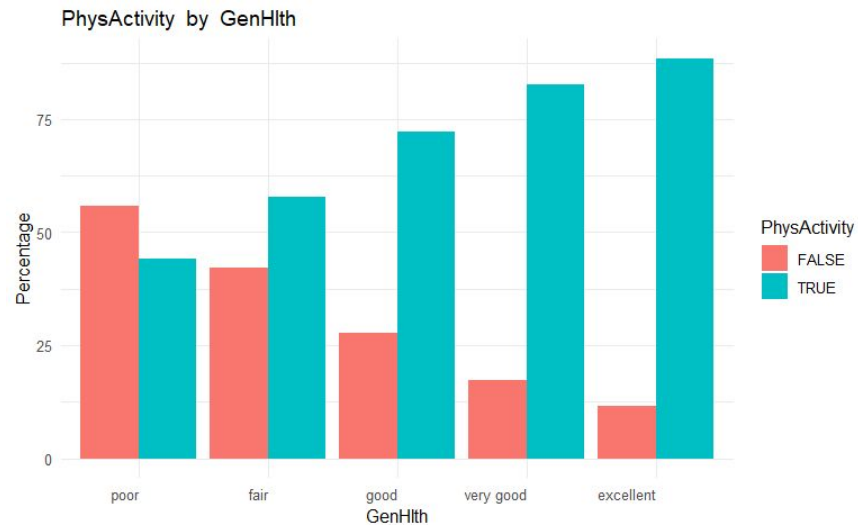
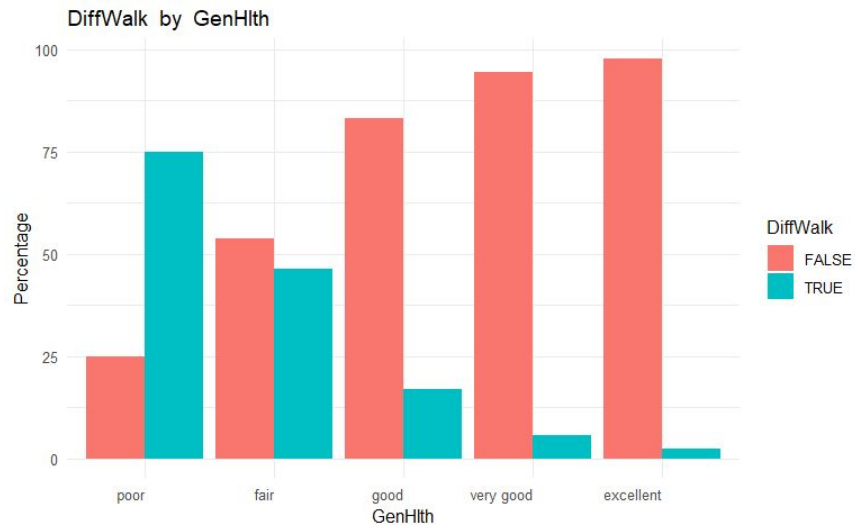


Залежність наявності діабету від суб'єктивної оцінки стану здоров'я.

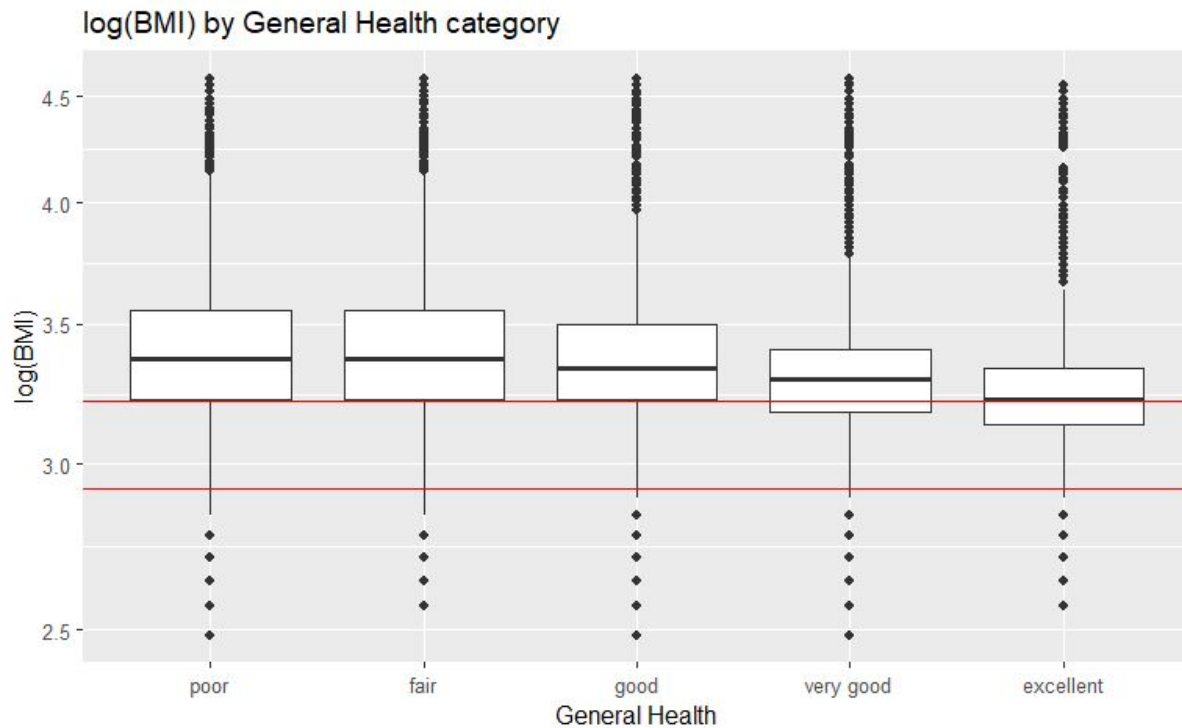
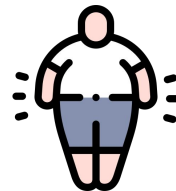
Чи завищують / занижують люди оцінку власного здоров'я (за відповідними критеріями)?



Чи завищують / занижують люди оцінку власного здоров'я
(за відповідними критеріями)?



Чи завищують / занижують люди оцінку власного здоров'я
(за відповідними критеріями)?

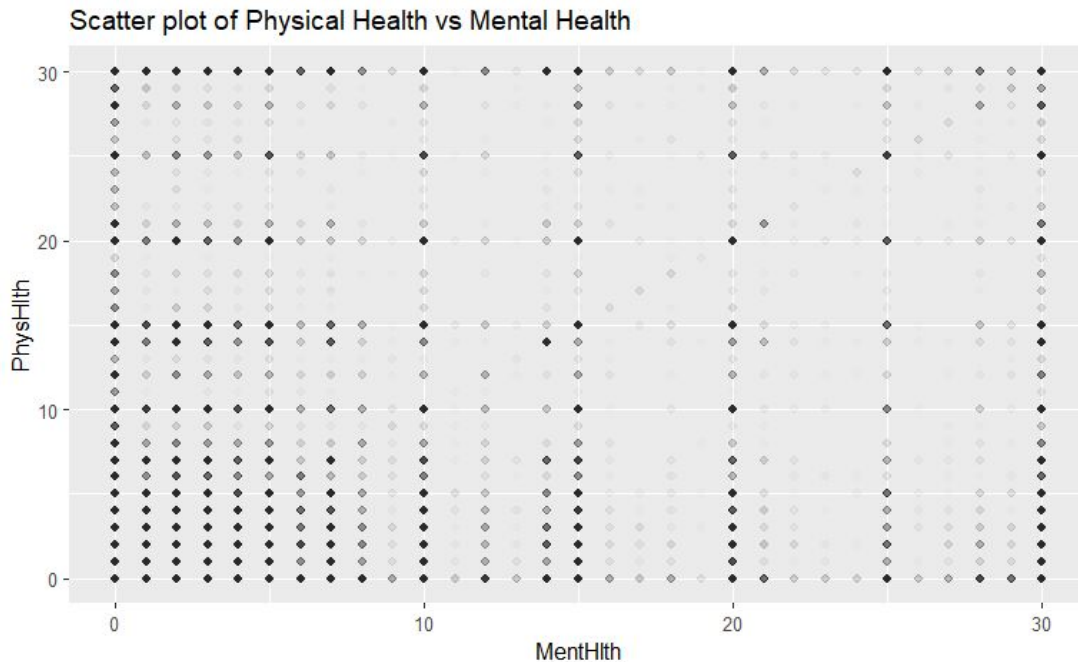




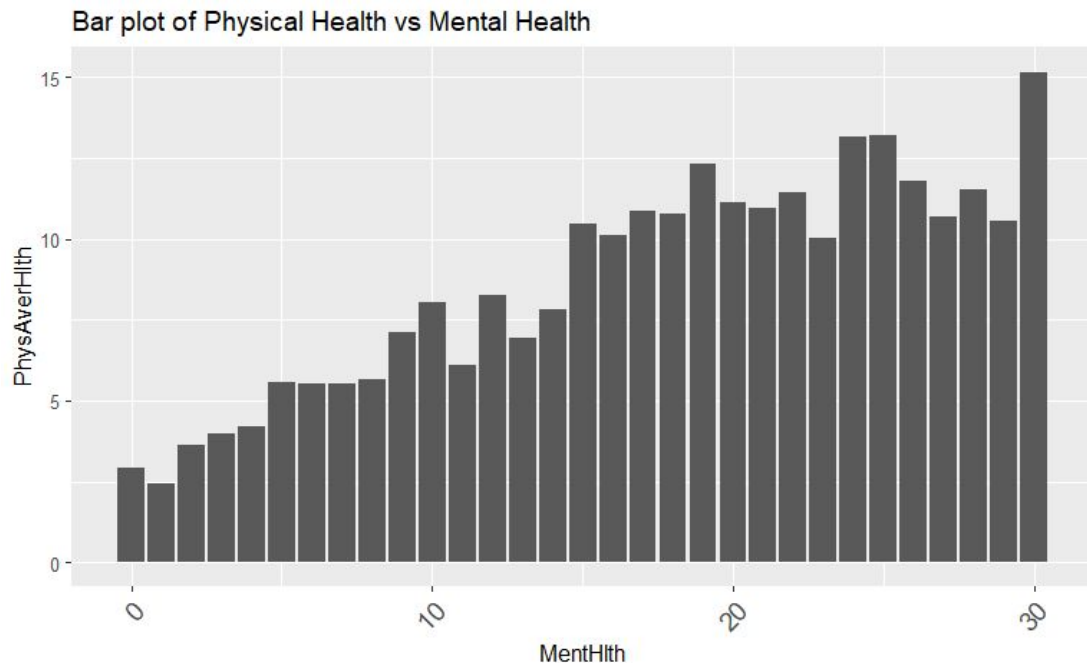
Питання 5



Чи впливає поганий ментальний стан на низький рівень фізичного здоров'я.



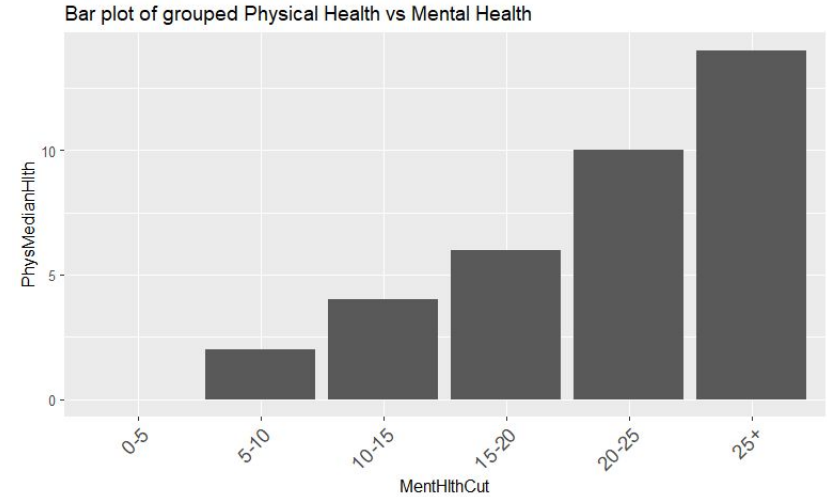
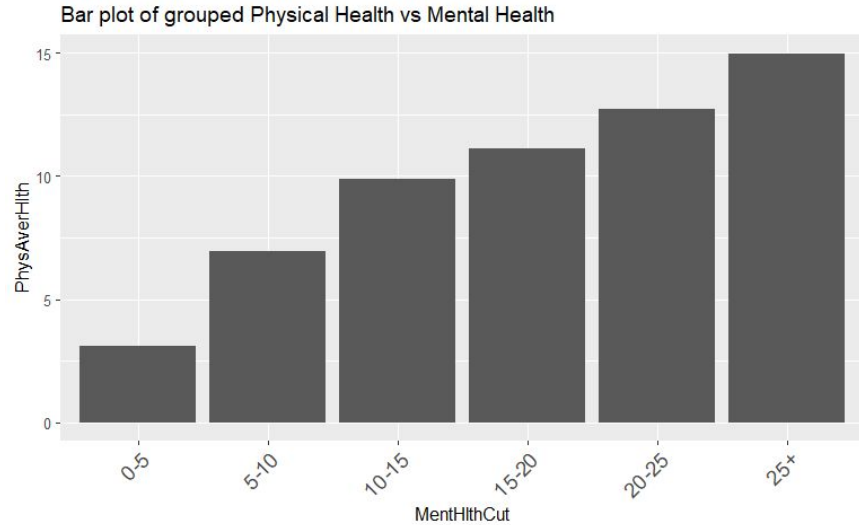
Залежність PhysHlth від MentHlth. (Середні вибірові PhysHlth в залежності від категорії MentHlth)



- Чим вище значення ментального здоров'я (чим більше днів людина мала погане ментальне здоров'я), тим вищий показник середнього PhysMeanHlth (кількість днів, коли у людини була травма, або людина була хвора, або почувала себе фізично погано).

Залежність PhysHlth від MentHlth.

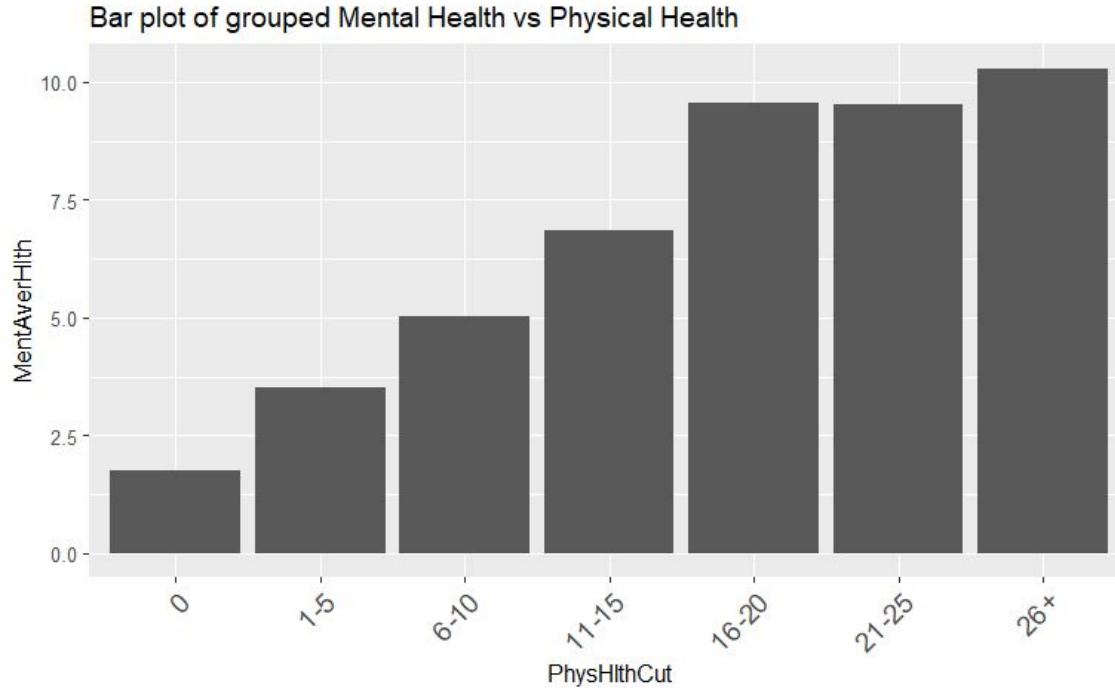
(Середні вибірові\медіани в залежності від категорії MentHlth)



- Має місце залежність. Чим вища категорія MentHlthCut - тим більше середнє\медіанне значення PhysHlth. Тобто чим гірше ментальне здоров'я - тим гірше фізичне.

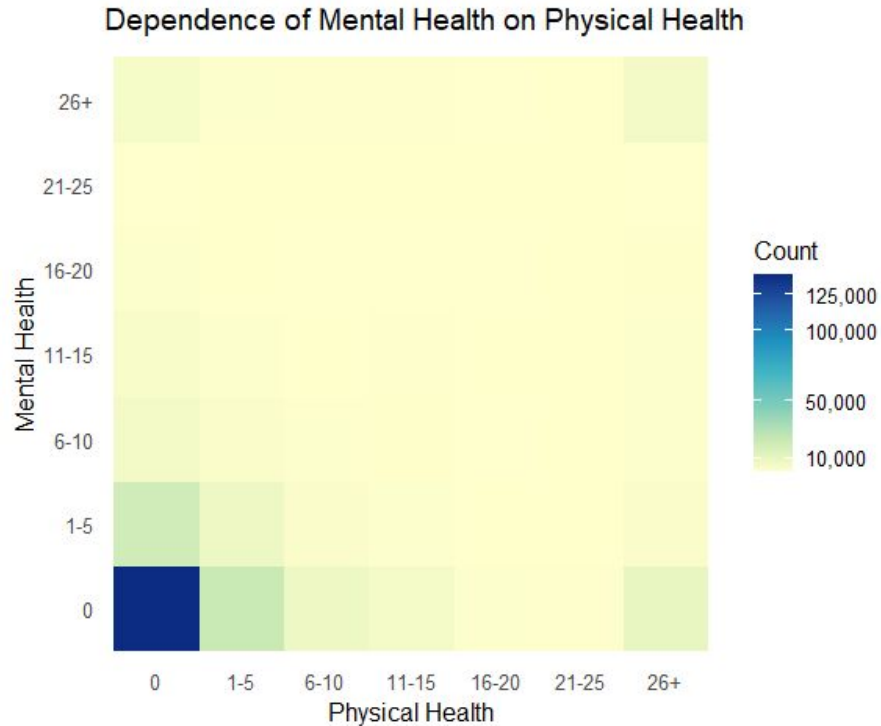
Залежність MentHlth від PhysHlthCut.

(Середні вибірові\медіани в залежності від категорії PhysHlth)



- Як можна побачити, якщо судити по вибірових середніх - залежність має місце.

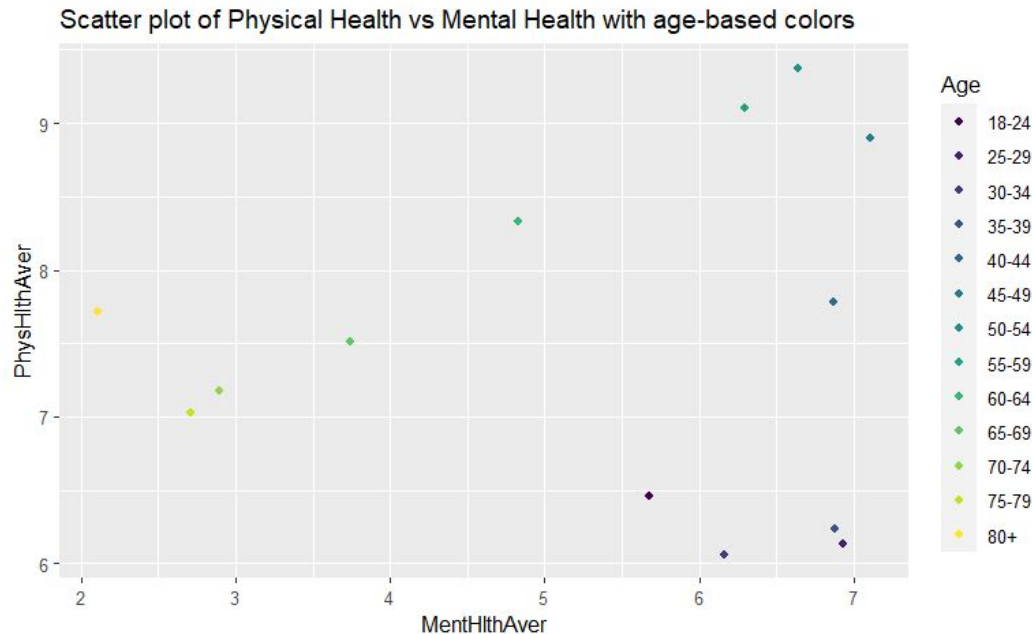
Теплова карта для згрупованих MentHlth та PhysHlth



- Найбільшою є частка людей з комбінації категорій "0 x 0".
- Також обширними категоріями є "1-5 x 0", "0 x 1-5", кутові комбінації "26+ x 0", "26+ x 26+", "0 x 26+".

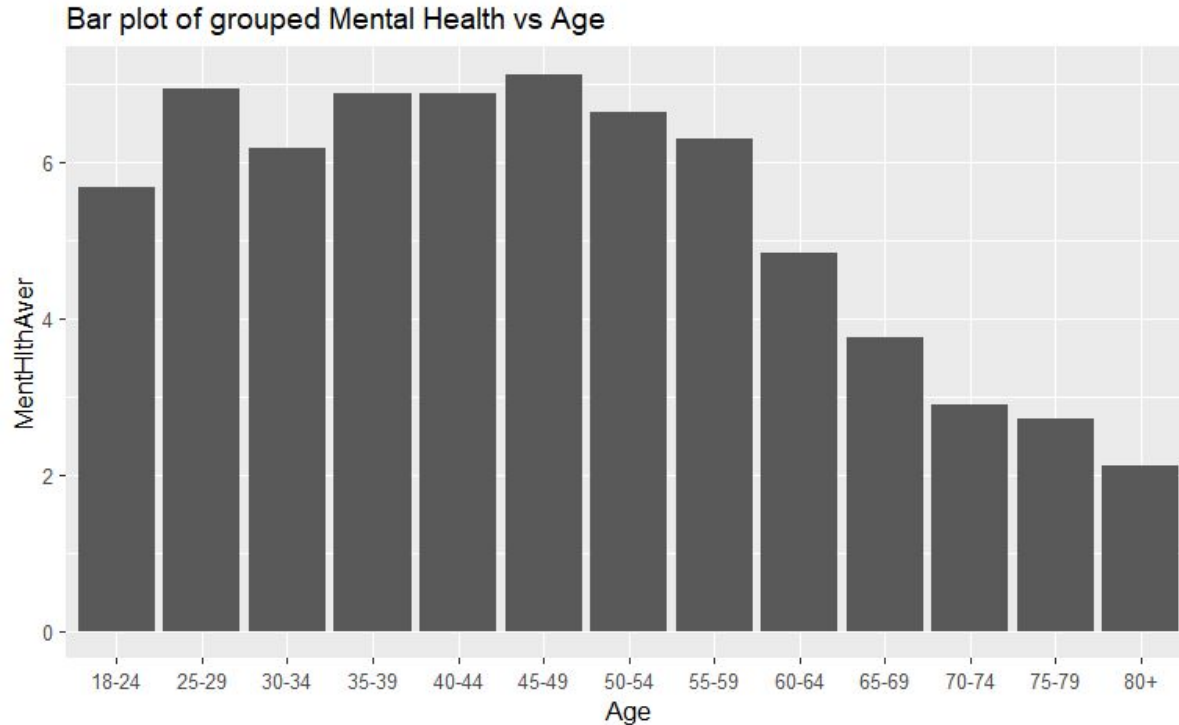
Питання 6

Залежність ментального здоров'я, оцінки фізичного здоров'я, від віку серед людей, які хворіють на діабет.



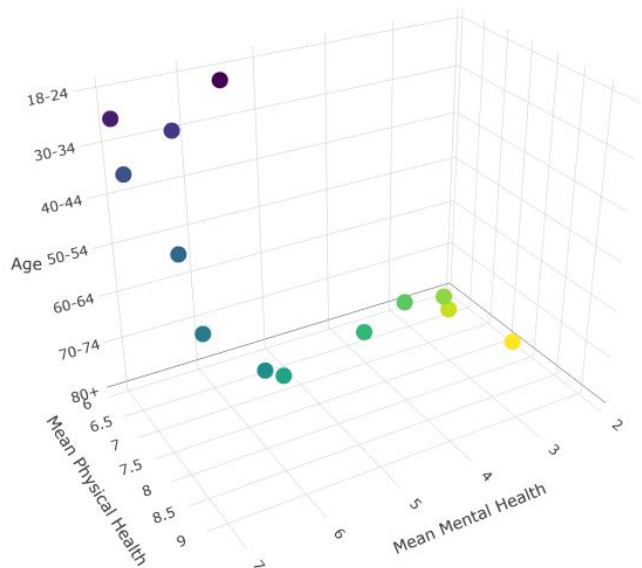
Діаграми розсіювання в залежності від вікової категорії для виб.середніх/медіан змінних PhysHlth і MentHlth серед людей хворих на діабет.

Стовпчикова діаграма в залежності від вікової категорії для медіан змінних PhysHlth і MentHlth серед людей хворих на діабет.

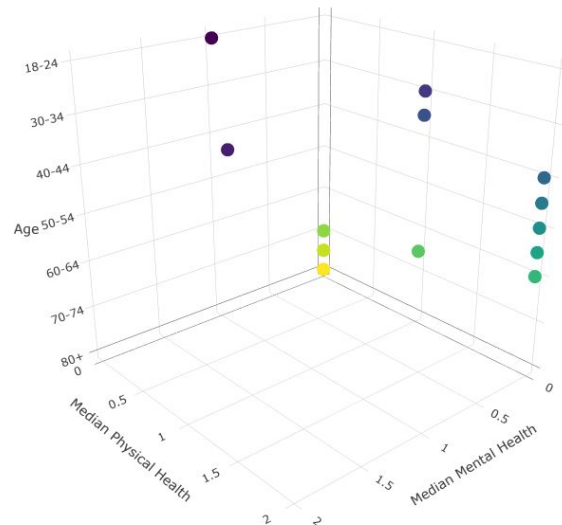


- Починаючи з вікової категорії 55-59 медіанне значення MentHlth спадає.

3-вимірні графіки діаграм розсіювання в залежності від вікової категорії для виб. середніх/медіан змінних PhysHlth і MentHlth серед людей хворих на діабет.



для вибірових середніх



для медіан

Висновки

- наявність діабету можуть впливати вік, ІМТ, деякі інші хвороби і навіть стать
- має місце вплив доходів та рівня освіти, а також видно, що діабетові сприяють вживання алкоголю, паління та погане харчування
- Третє питання: люди без медичного страхування рідше обстежуються на рівень холестерину; додаткові питання
 - чи є залежність наявності страховки від заробітної плати?
 - чи є залежність змінної NoDocbcCost від заробітної плати?
 - залежність наявності діабету від змінних CholCheck, NoDocbcCost
- видно залежність страховки від доходів, і залежність наявності діабету від обстеження холестерину та труднощів із вибором лікаря
- спостерігається зменшення частки людей з діабетом по мірі покращення оцінки свого стану
- бачимо покращення рівня здоров'я при підвищенні оцінки свого здоров'я, зокрема, - ІМТ

Висновки

- є певний зв'язок середніх показників ментального та фізичного здоров'я, однак важко стверджувати, що по одному параметру можна судити про інший
- спостерігається погіршення середнього показника ментального по мірі збільшення віку опитуваних, тоді як фізичне здоров'я майже не залежить від віку

Дякую за увагу!

Джерела:

1. [Introduction to ggridges](#). (дата звернення: 28.05.2024)
2. [3D Scatter Plots in R. URL](#).
3. Ідеї для побудови групованих стовпчикових діаграм та “Violin” графіків були взяті з [галереї графіків в R](#).
4. Матеріали лекції 2.