

Лабораторна робота №1 «Розвідковий аналіз даних»

Виконали ст. групи км-11:

Команда “CrAzy_pEnGuIn_cOdErs”

Шушпаннікова Інна

Кракович Борислав

Ягода Євгенія

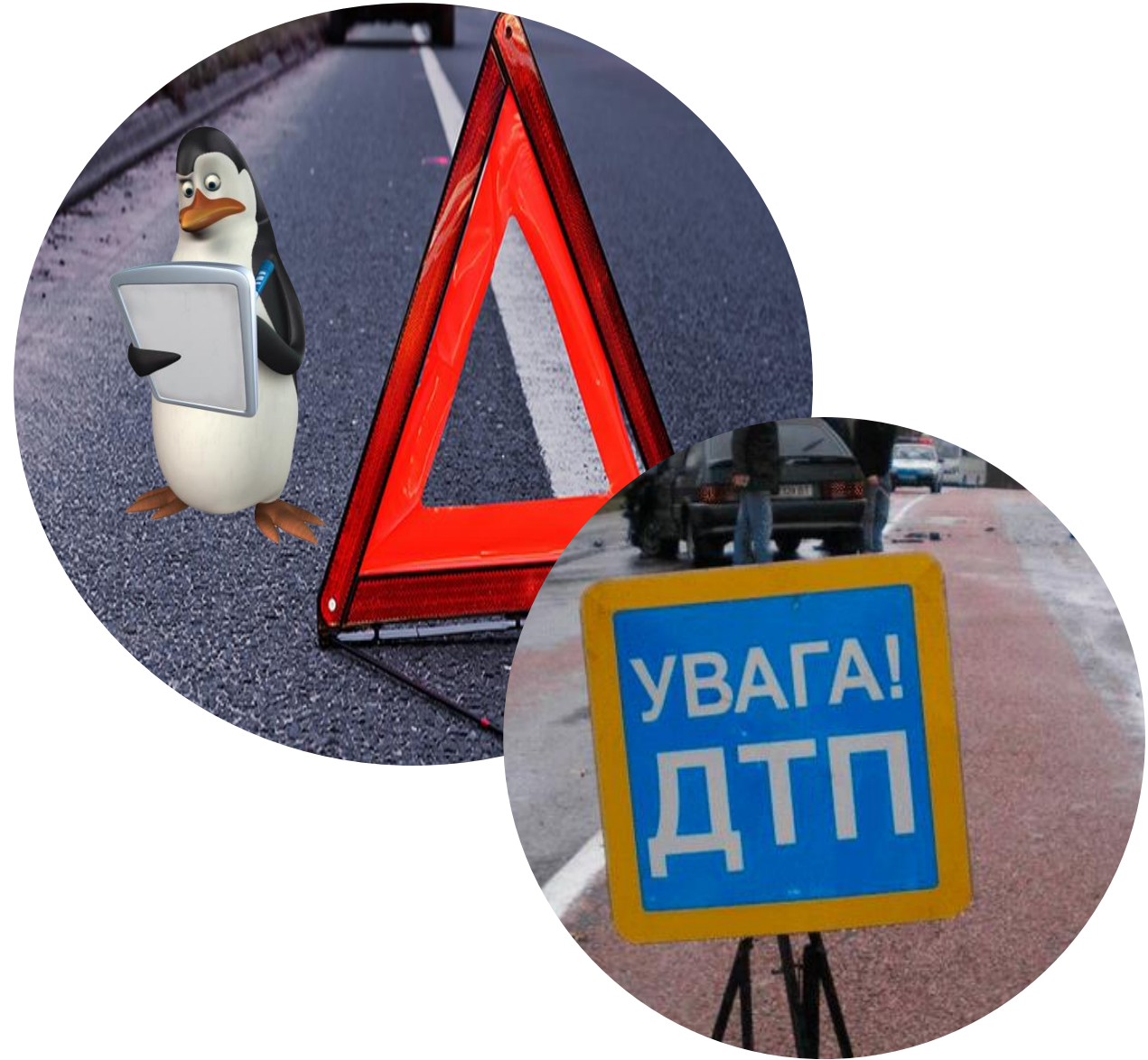
Лушников Іван

Атаман Юлія



Вступ

- Дорожньо-транспортні пригоди становлять серйозну проблему, яка залишає свій слід на житті та здоров'ї багатьох людей щодня. Кожного року на дорогах по всьому світу трапляється безліч інцидентів, які призводять до трагічних наслідків та значних матеріальних збитків. Статистика свідчить про те, що це проблема, яка потребує невідкладного вирішення.
- У світлі цих викликів наша команда вирішила зосередитися на дослідженні факторів, що впливають на виникнення та наслідки дорожньо-транспортних пригод. Наша мета полягала в тому, щоб розкрити та проаналізувати ці аспекти, щоб виявити можливі шляхи покращення безпеки на дорогах та зменшення кількості інцидентів.



ОПИС ДАНИХ

- Датасет: **UK Road Safety: Traffic Accidents and Vehicles**
- Уряд Великобританії збирає та публікує (зазвичай щорічно) детальну інформацію про дорожньо-транспортні пригоди по всій країні. Ця інформація включає географічне розташування, погодні умови, типи транспортних засобів, кількість постраждалих і маневри транспортних засобів, що робить цей набір дуже цікавим і вичерпним для аналізу та дослідження.
- Датасет демонструє дорожньо-транспортні пригоди та залучені транспортні засоби у Великобританії (2005-2017 рр.). Він містить два файли:
 - Accident_Information.csv(2005-2017): розмір-2047256, 34 змінні. Надає дані про дорожньо-транспортну пригоду.
 - Vehicle_Information.csv(2004-2016): розмір-1488981, 24 змінні. Надає дані транспортного засобу та власника.



«Брудний» датасет має наступну структуру



Accident_Information

1. Accident_Index-Індекс_ДТП
2. Accident_Severity-Серйозність аварії
3. Carriageway Hazards -
Небезпека проїжджої частини
4. Date - Дата
5. Day_of_Week – День тижня
6. Junction Detail-перехрестя
7. Light_Conditions -
Світлові_умови
8. Number_of_Casualties-
Кількість_загиблих

9. Road Surface Conditions-Стан
дорожнього покриття
10. Road Type-Тип дороги
11. Speed limit-Обмеження
швидкості
12. Time-час
13. Urban or Rural_Area-Міська або
сільська місцевість
14. Weather Conditions-
Метеорологічні умови
15. Year-Рік

Vehicle_Information

1. Accident_Index- Індекс_ДТП
2. Age_Band_of_Driver-Вік
автомобіля 1,2
3. Age_of_Vehicle -Вік водія 1,2
4. Sex_of_Driver -Стать водія 1,2
5. make-Марка ТЗ 1,2
6. Vehicle_Manoeuvre -Маневр ТЗ
1,2 під час аварії
7. Vehicle_Type -Тип ТЗ 1,2
8. X1st_Point_of_Impact-Точка
зіткнення відносно даного ТЗ
1,2

Опрацювання даних

1. Обираємо 2 транспортні засоби та дані з 2006-2015 роки
2. Видаляємо індекси яких немає у 2 файлі, але є у першому(навпаки)
3. Відбираємо парні та непарні рядки в Vehicle і додаємо в df2 та df1 відповідно
4. Перейменовуємо змінні в df2 та df1 на : Car_Brand1=make, Point_of_Impact1=X1st_Point_of_Impact
5. Об'єднуємо файли за колонкою Accident_Index s_df1, s_df2
6. Об'єднуємо файли s_Accident_Index_filtered , merged_Vehicle за індексом аварії
7. Видаляємо з колонок значення; na, Data missing or out of range, Unknown, Not known
8. Додаємо дві колонки day та month розділяючи та видаляючи колонку Date
9. Отримуємо датафрейм filtered_data_3 = 210526 записів та 30 змінних.

NA

Age_of_Vehicle1 – 16%

Age_of_Vehicle2 – 18%

Data missing or out of range

Road_Surface_Conditions – 0,13%

Age_Band_of_Driver1 – 5,65%

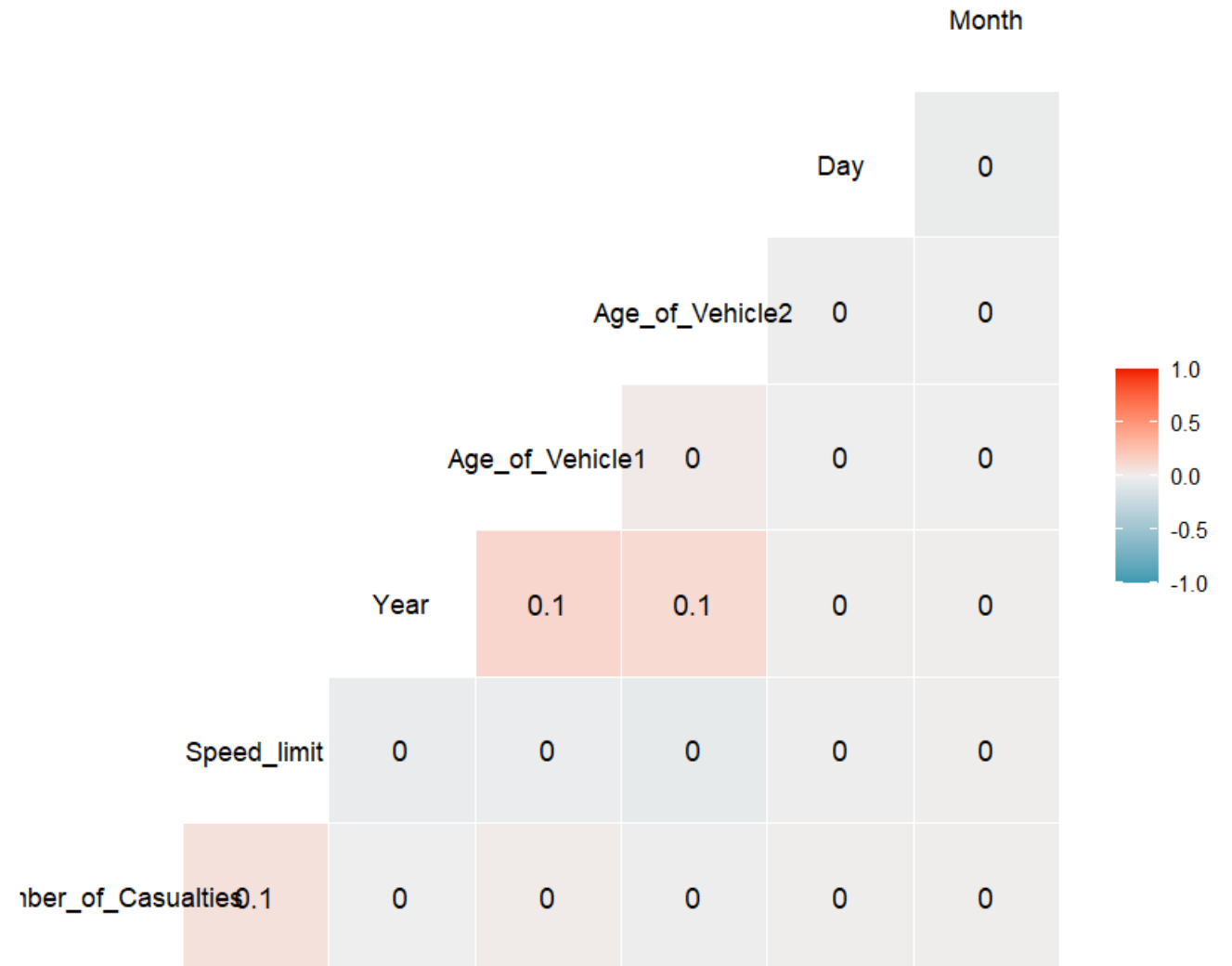
Age_Band_of_Driver1 – 7,59%



Загальна структура датафрейму після опрацювання

```
'data.frame': 210526 obs. of 30 variables:
 $ Accident_Index      : chr  "200601BS70001" "200601BS70046" "200601BS70064" "200601BS70076" ...
 $ Accident_Severity   : chr  "Slight" "Slight" "Serious" "Slight" ...
 $ Carriageway_Hazards : chr  "None" "None" "None" "None" ...
 $ Day_of_Week         : chr  "Wednesday" "Wednesday" "Wednesday" "Tuesday" ...
 $ Junction_Detail     : chr  "T or staggered junction" "Crossroads" "T or staggered junction" "Crossroads" ...
 $ Light_Conditions    : chr  "Daylight" "Daylight" "Daylight" "Darkness - lights lit" ...
 $ Number_of_Casualties : int  1 1 1 1 2 1 1 1 1 1 ...
 $ Road_Surface_Conditions: chr  "Dry" "Dry" "Wet or damp" "Wet or damp" ...
 $ Road_Type          : chr  "Single carriageway" "Single carriageway" "Single carriageway" "Single carriageway" ...
 $ Speed_limit         : int  30 30 30 30 30 30 30 30 30 30 ...
 $ Area_Type          : chr  "Urban" "Urban" "Urban" "Urban" ...
 $ Weather_Conditions  : chr  "Fine no high winds" "Fine no high winds" "Snowing no high winds" "Raining no high winds" ...
 $ Year               : int  2006 2006 2006 2006 2006 2006 2006 2006 2006 2006 ...
 $ Time               : chr  "15:40" "16:54" "11:48" "23:49" ...
 $ Age_Band_of_Driver1 : chr  "26 - 35" "36 - 45" "26 - 35" "26 - 35" ...
 $ Age_of_Vehicle1    : int  12 4 13 7 2 8 7 1 1 11 ...
 $ Sex_of_Driver1     : chr  "Male" "Male" "Female" "Male" ...
 $ Car_Brand1         : chr  "HONDA" "MAZDA" "FORD" "FIAT" ...
 $ Vehicle_Manoeuvre1 : chr  "Going ahead other" "Waiting to turn right" "Going ahead other" "Turning right" ...
 $ Vehicle_Type1      : chr  "Motorcycle over 125cc and up to 500cc" "Car" "Car" "Car" ...
 $ Point_of_Impact1   : chr  "Front" "Offside" "Front" "Front" ...
 $ Age_Band_of_Driver2 : chr  "36 - 45" "26 - 35" "46 - 55" "26 - 35" ...
 $ Age_of_Vehicle2    : int  10 1 2 3 8 4 7 1 4 1 ...
 $ Sex_of_Driver2     : chr  "Male" "Female" "Female" "Male" ...
 $ Car_Brand2         : chr  "MITSUBISHI" "HONDA" "MERCEDES" "PIAGGIO" ...
 $ Vehicle_Manoeuvre2 : chr  "Turning right" "Going ahead other" "Waiting to go - held up" "Going ahead other" ...
 $ Vehicle_Type2      : chr  "Car" "Motorcycle 125cc and under" "Car" "Motorcycle 125cc and under" ...
 $ Point_of_Impact2   : chr  "Front" "Front" "Back" "Nearside" ...
 $ Day               : int  18 8 22 28 13 28 30 28 23 9 ...
 $ Month             : int  1 2 2 2 1 3 3 4 4 5 ...
```

Корелограма числових даних датасету



Дослідницькі питання

01

Як особливості місцевості та дорожнього покриття впливають на ймовірність виникнення аварій та тяжкість?

02

Як впливають характеристики автомобіля на кількість/серйозність аварій?

03

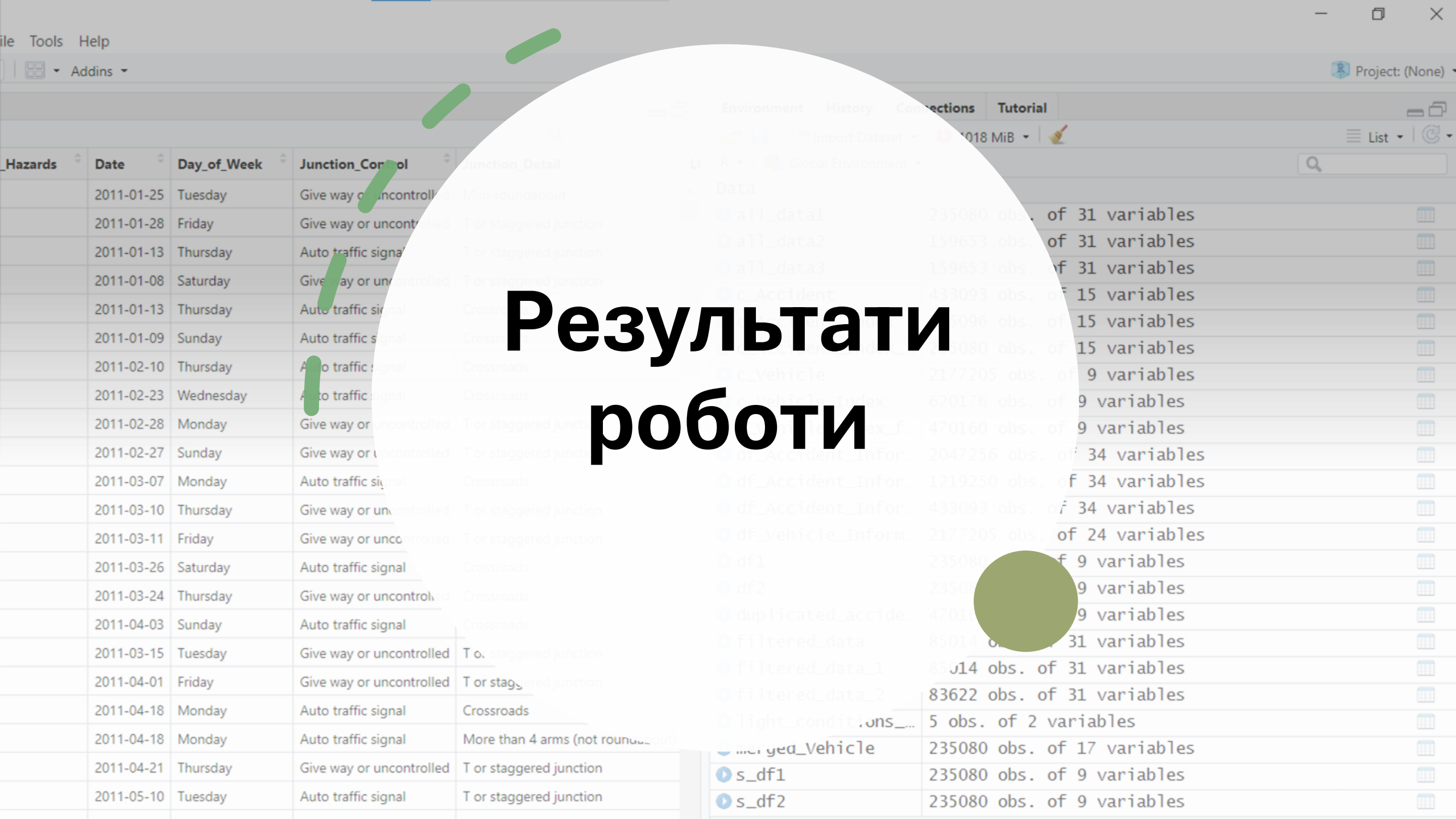
Як впливає час/день/рік на кількість/серйозність аварій?

04

Вплив фізичних особливостей водія на поведінку на дорозі?

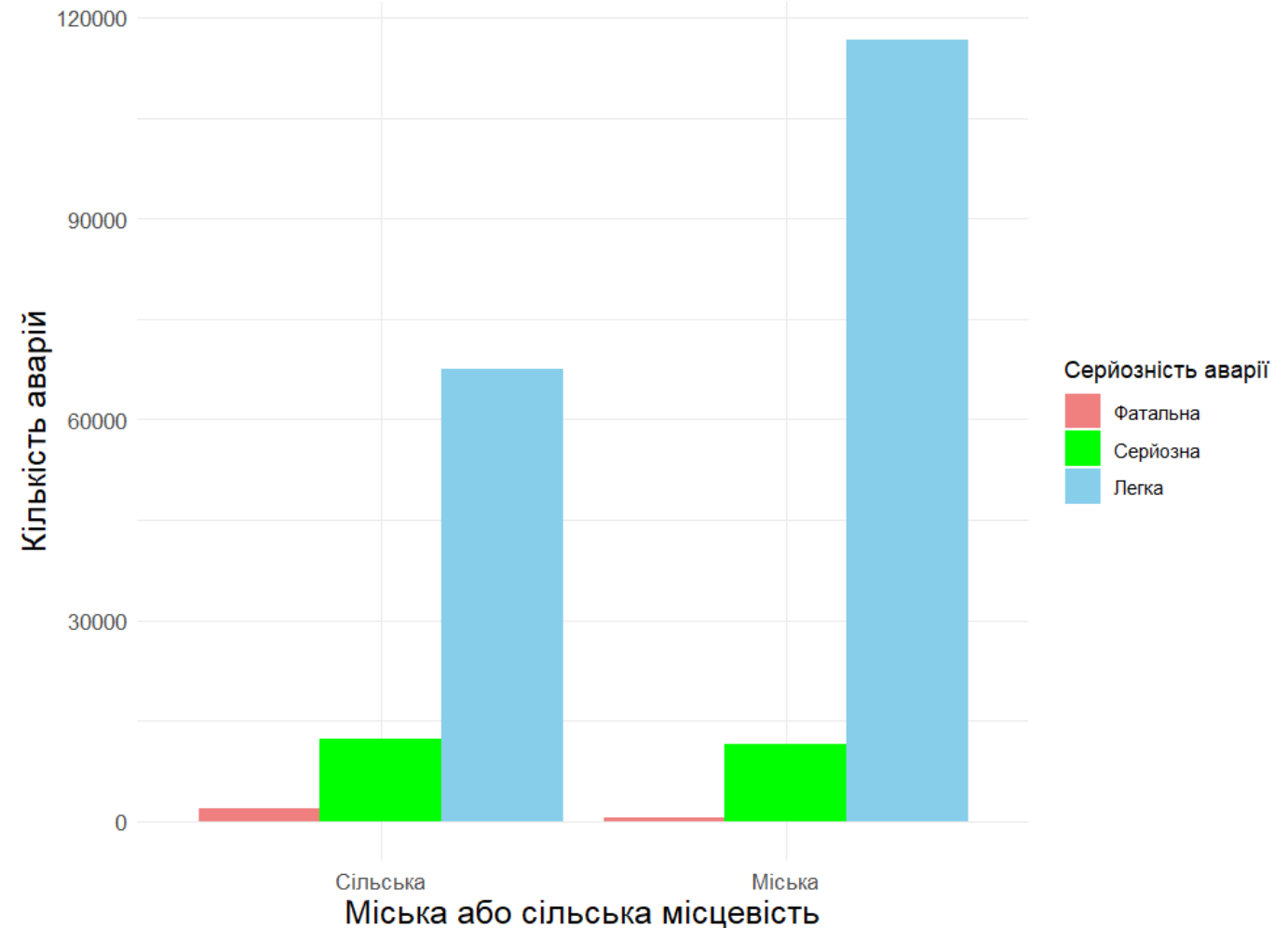
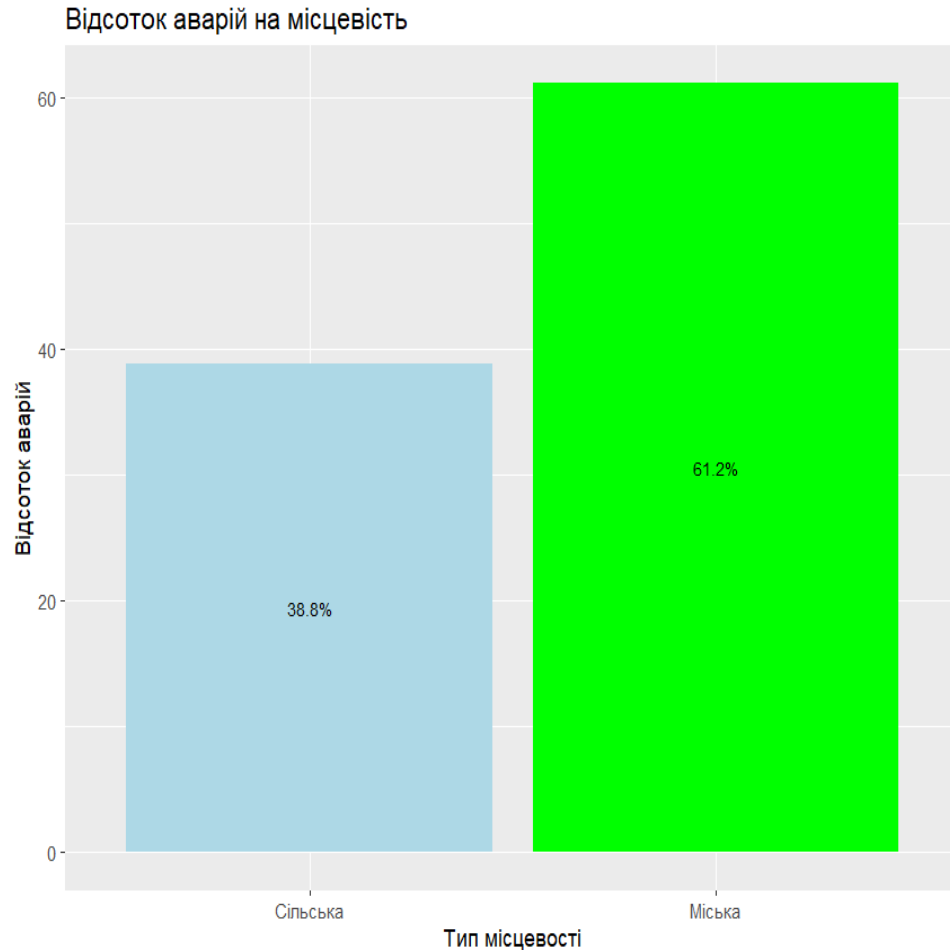
05

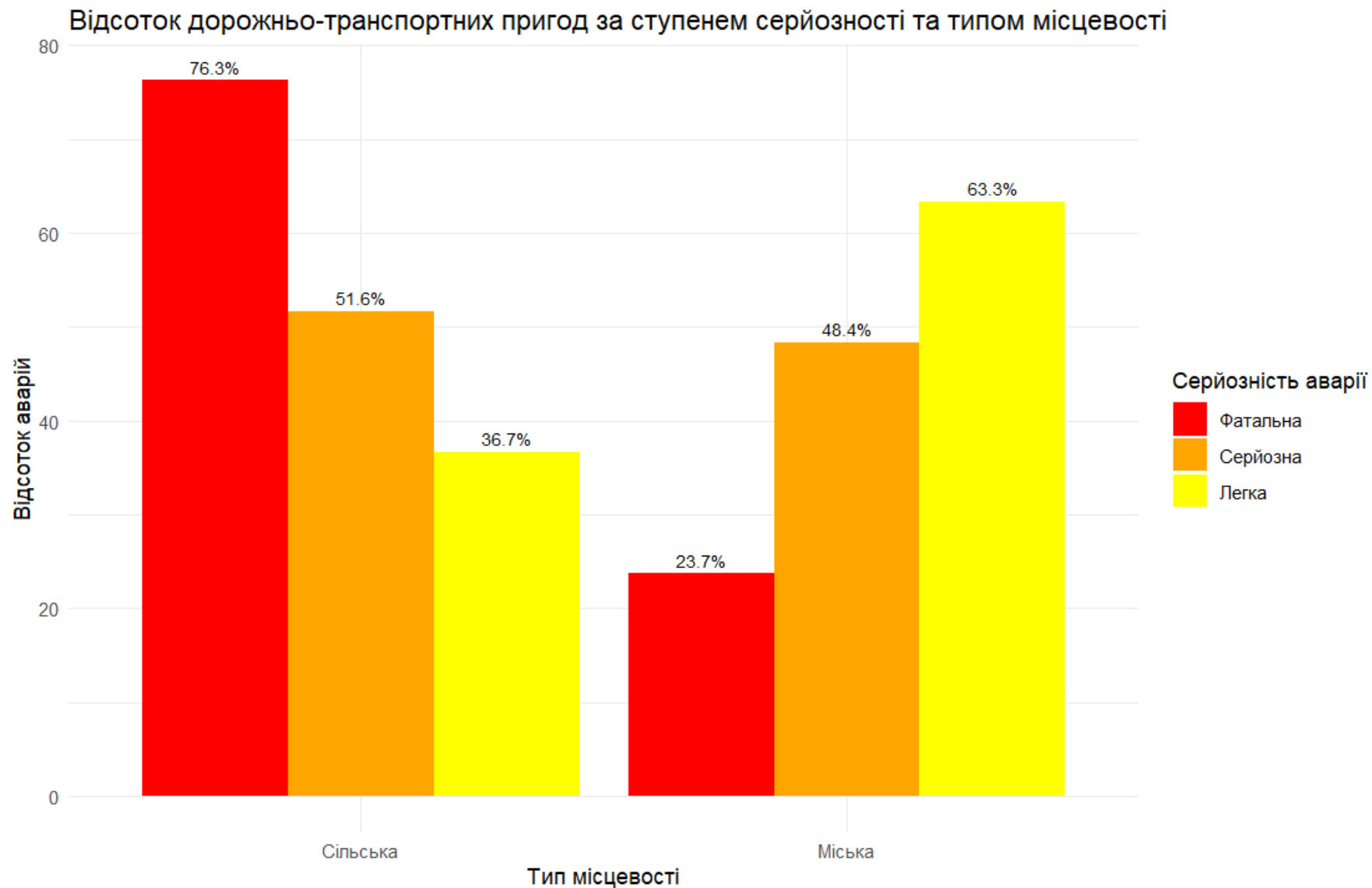
Як фактори на які не може впливати водій впливають на кількість/серйозність аварій?



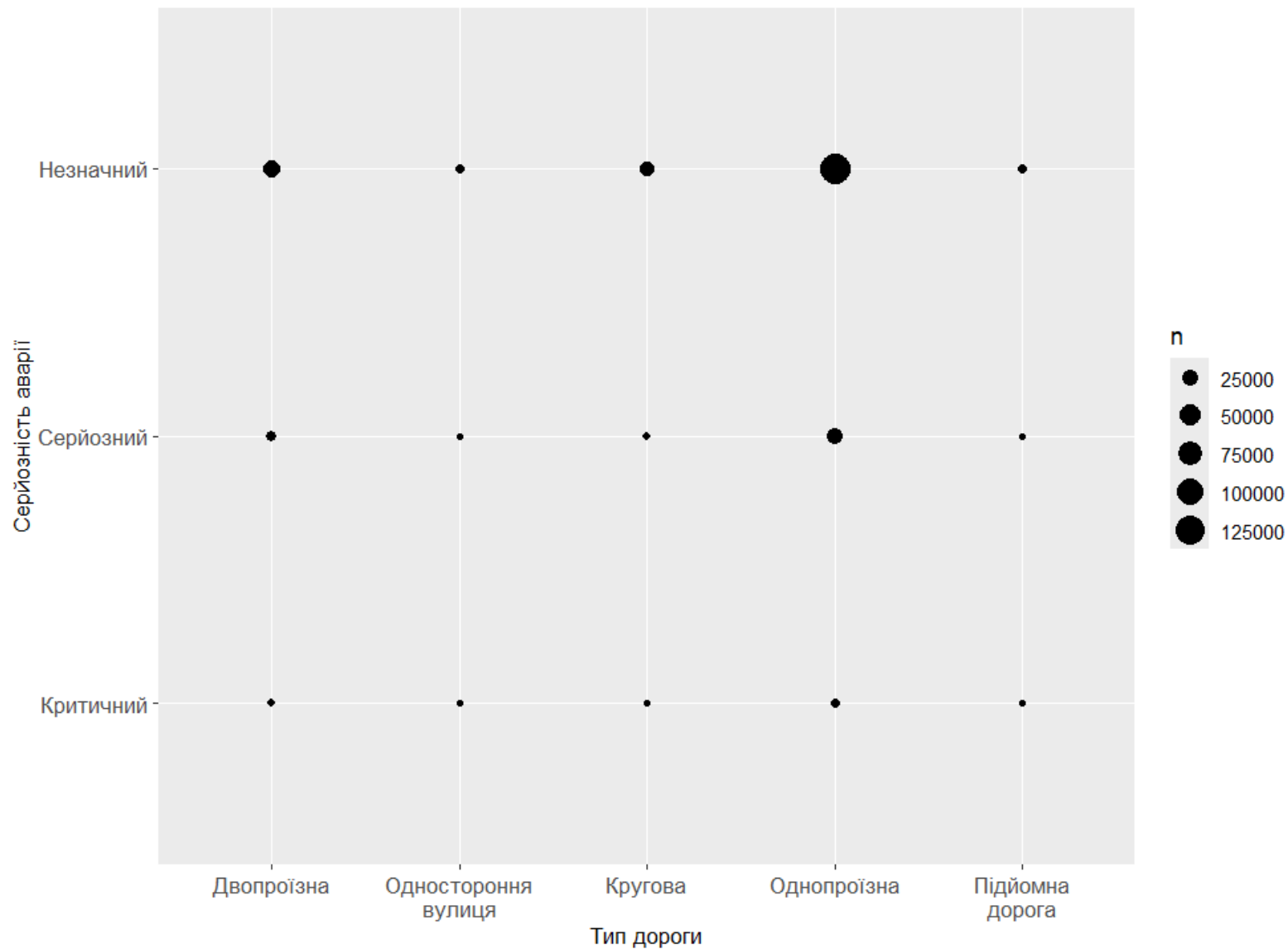
Результати
роботи

1. Як особливості місцевості та дорожнього покриття впливають на ймовірність виникнення аварій та тяжкість?





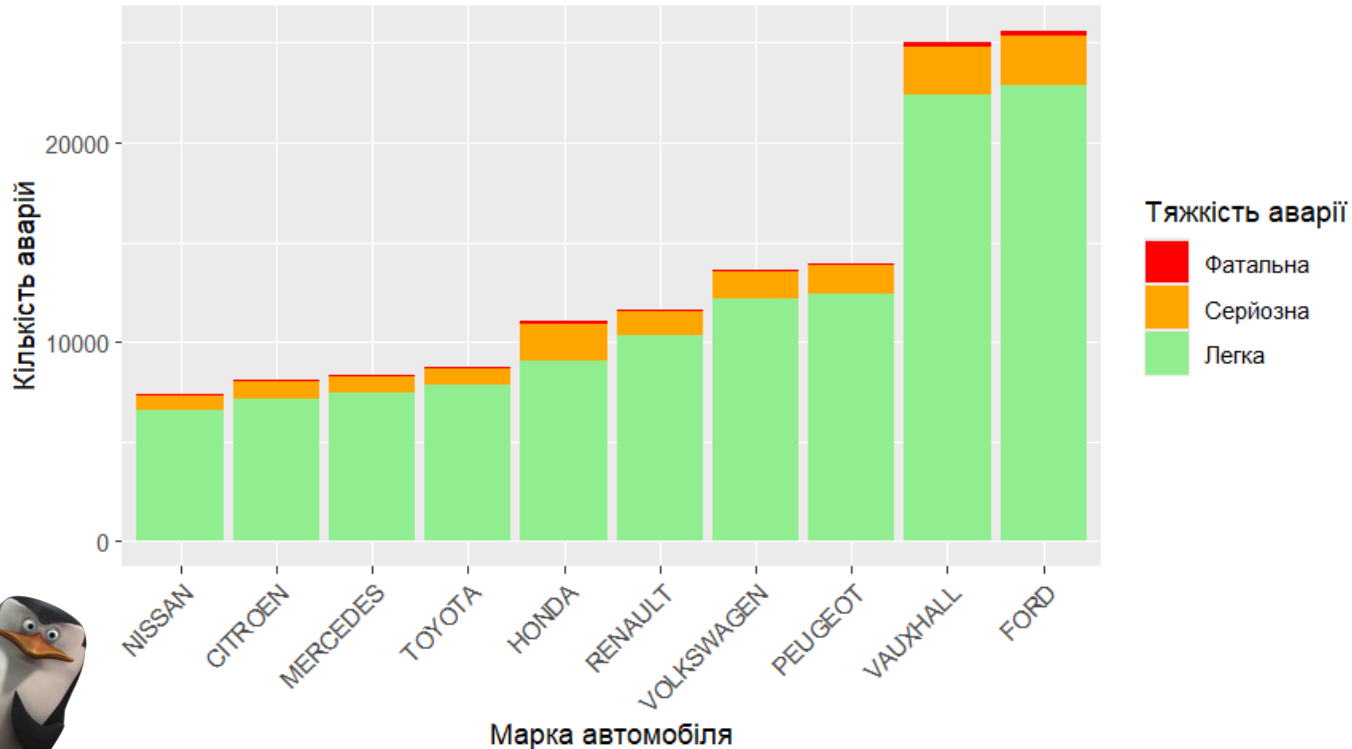
Як бачимо, в сільській місцевості аварії трапляються не так часто, проте частка фатальних ДТП набагато вища, ніж у міській місцевості.



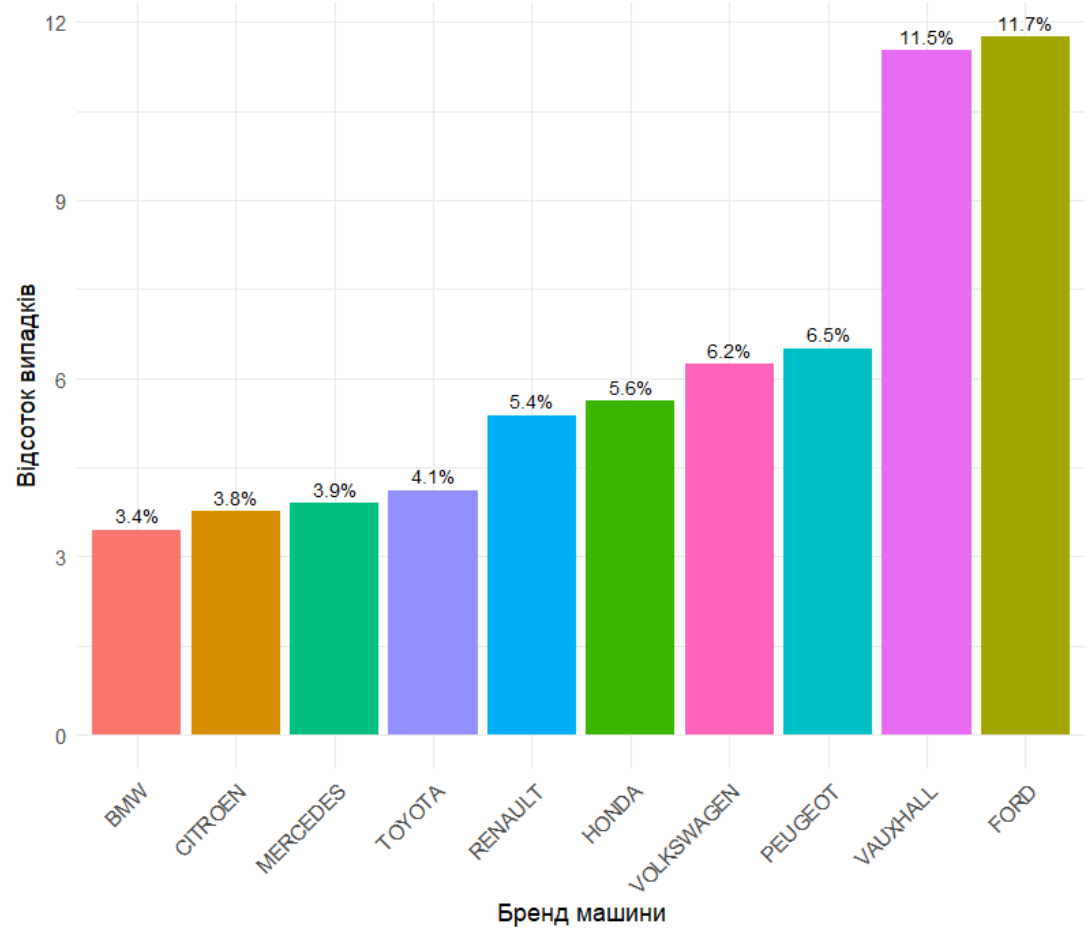
- Найбільша кількість незначних аварій відбувається на однопроїзному та двопроїзному типі дороги
- Також досить серйозні та критичні аварії за ступенем тяжкості трапляються частіше.

2. Як впливають характеристики автомобіля на кількість/серйозність аварій?

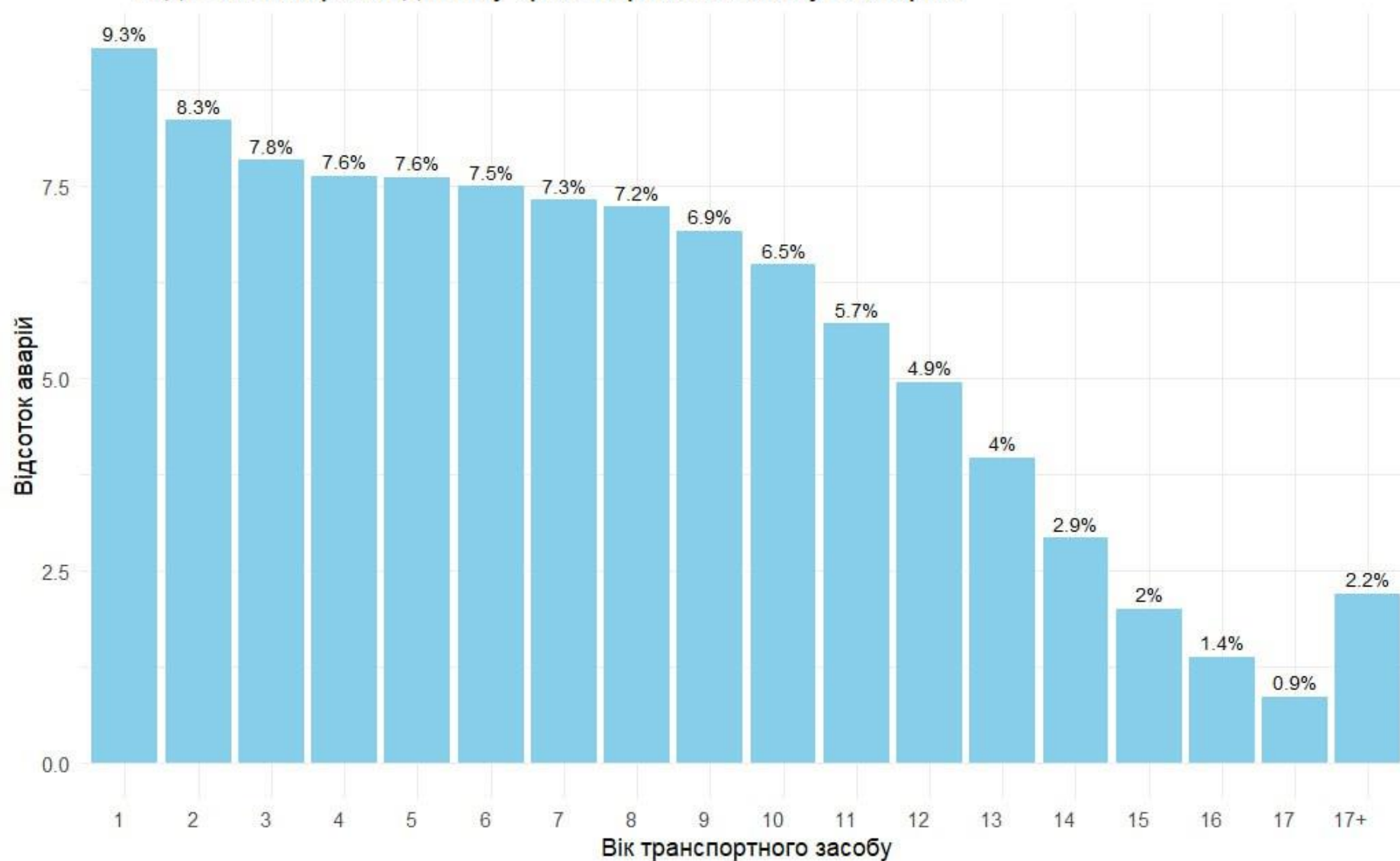
Топ-10 найпопулярніших марок автомобілів із найбільшою кількістю аварій



Десятка найпопулярніших марок автомобілів

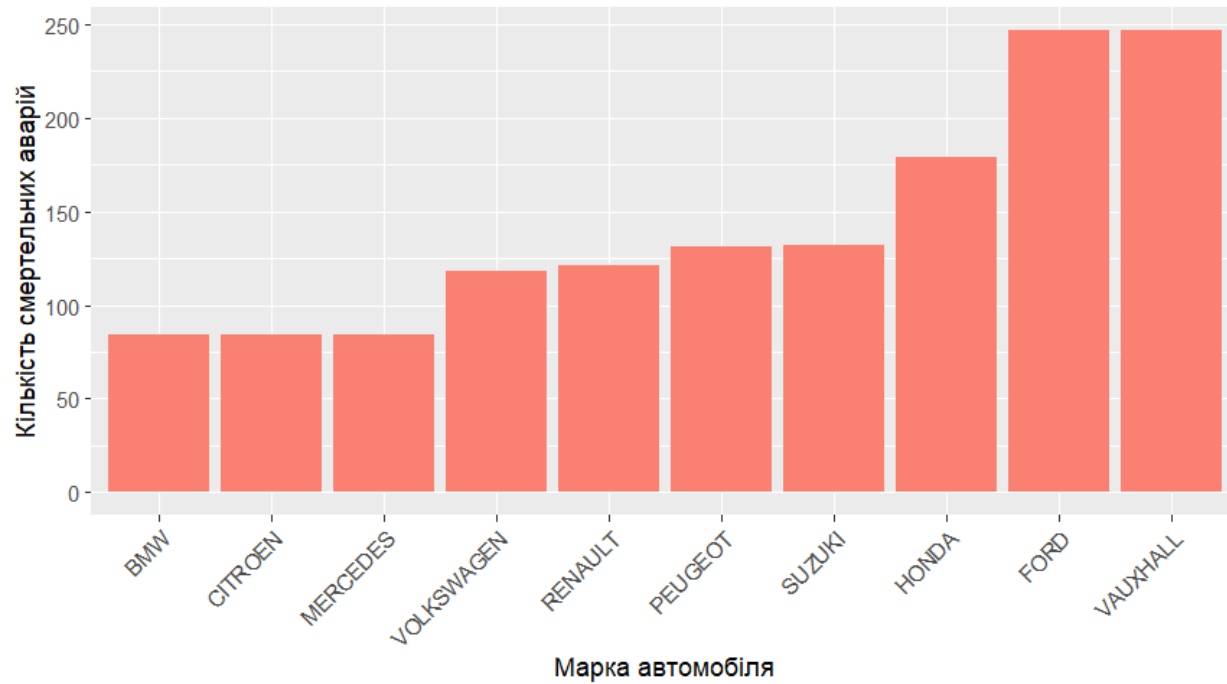


Відсотковий розподіл віку транспортного засобу в аваріях



Найбільша частка аварій близько 9,3% у автомобілів до 1 року використання

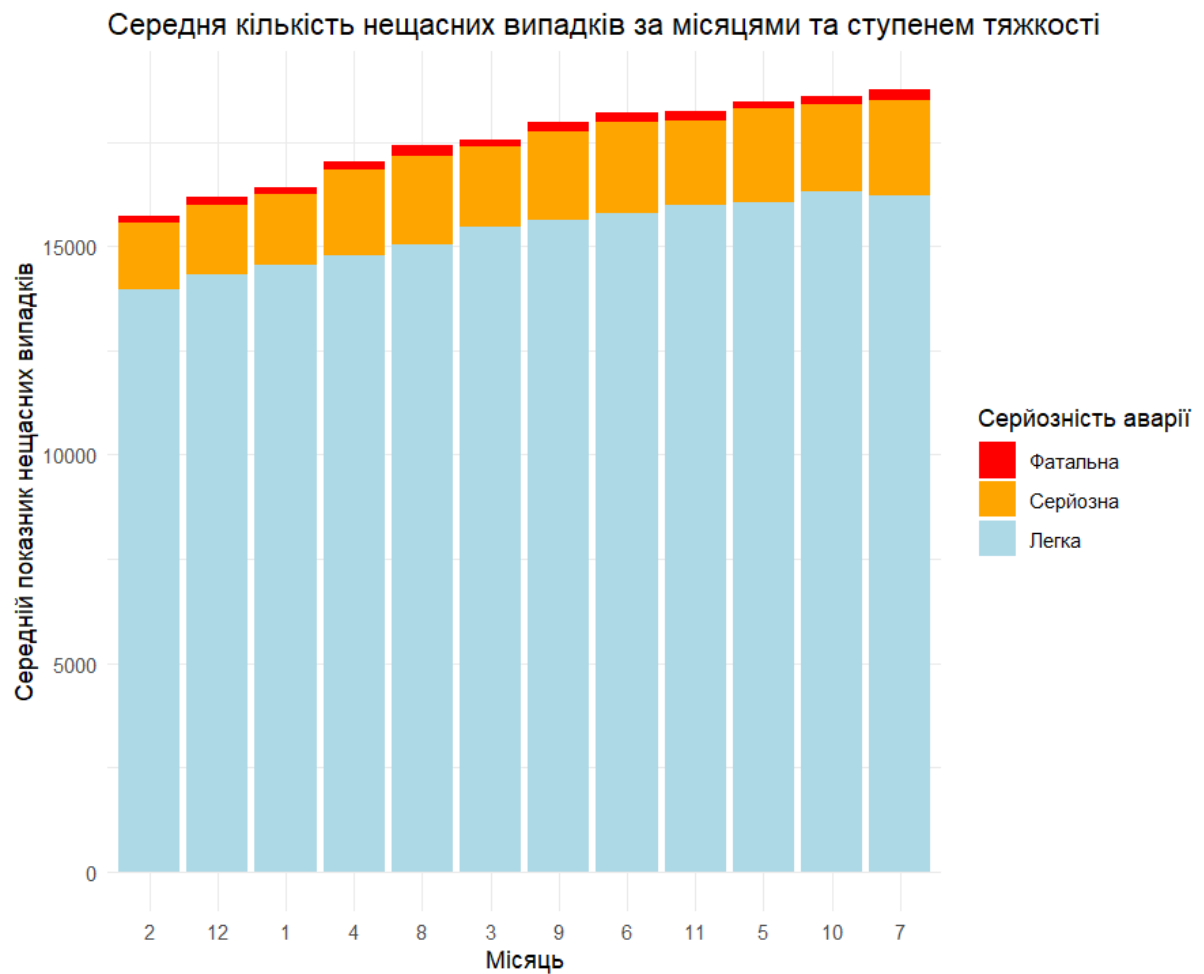
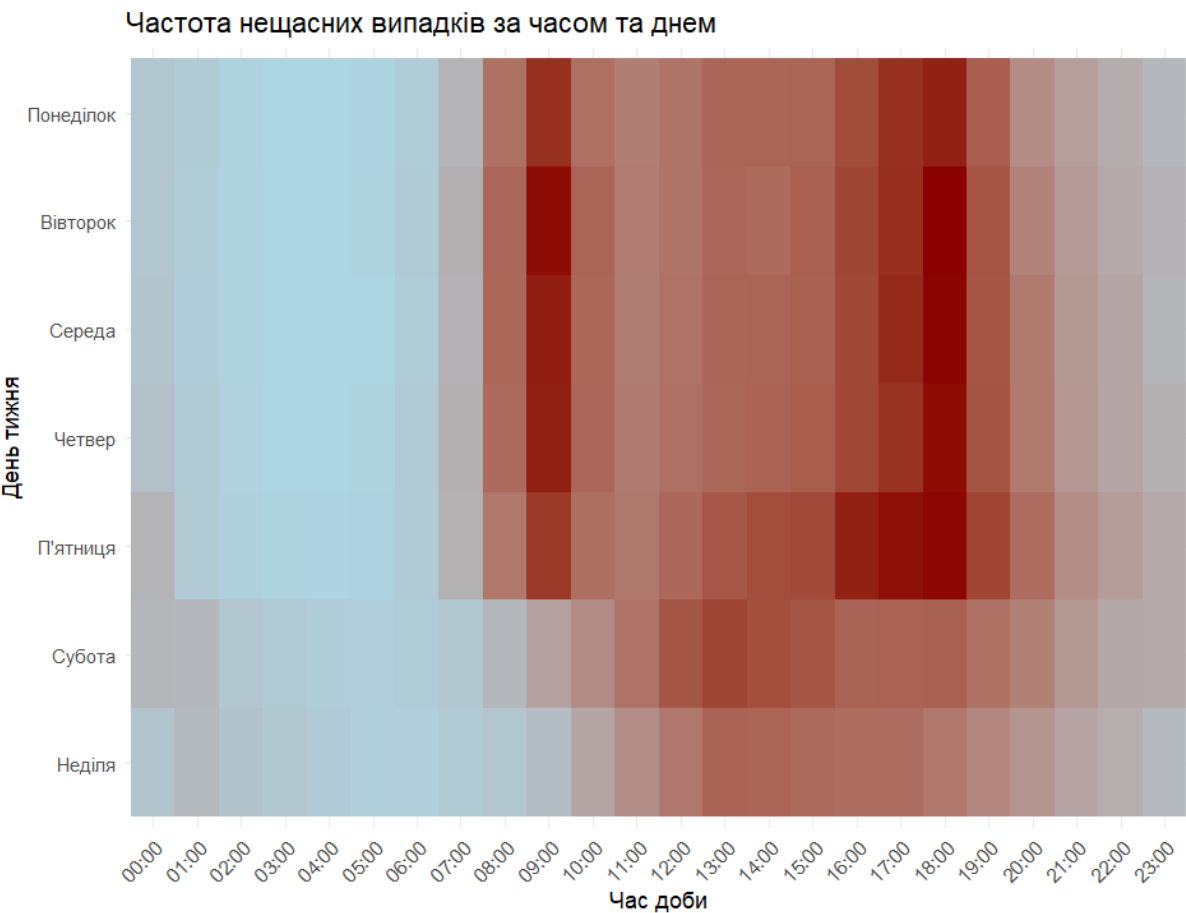
Топ-10 автомобільних марок
за кількістю смертельних аварій

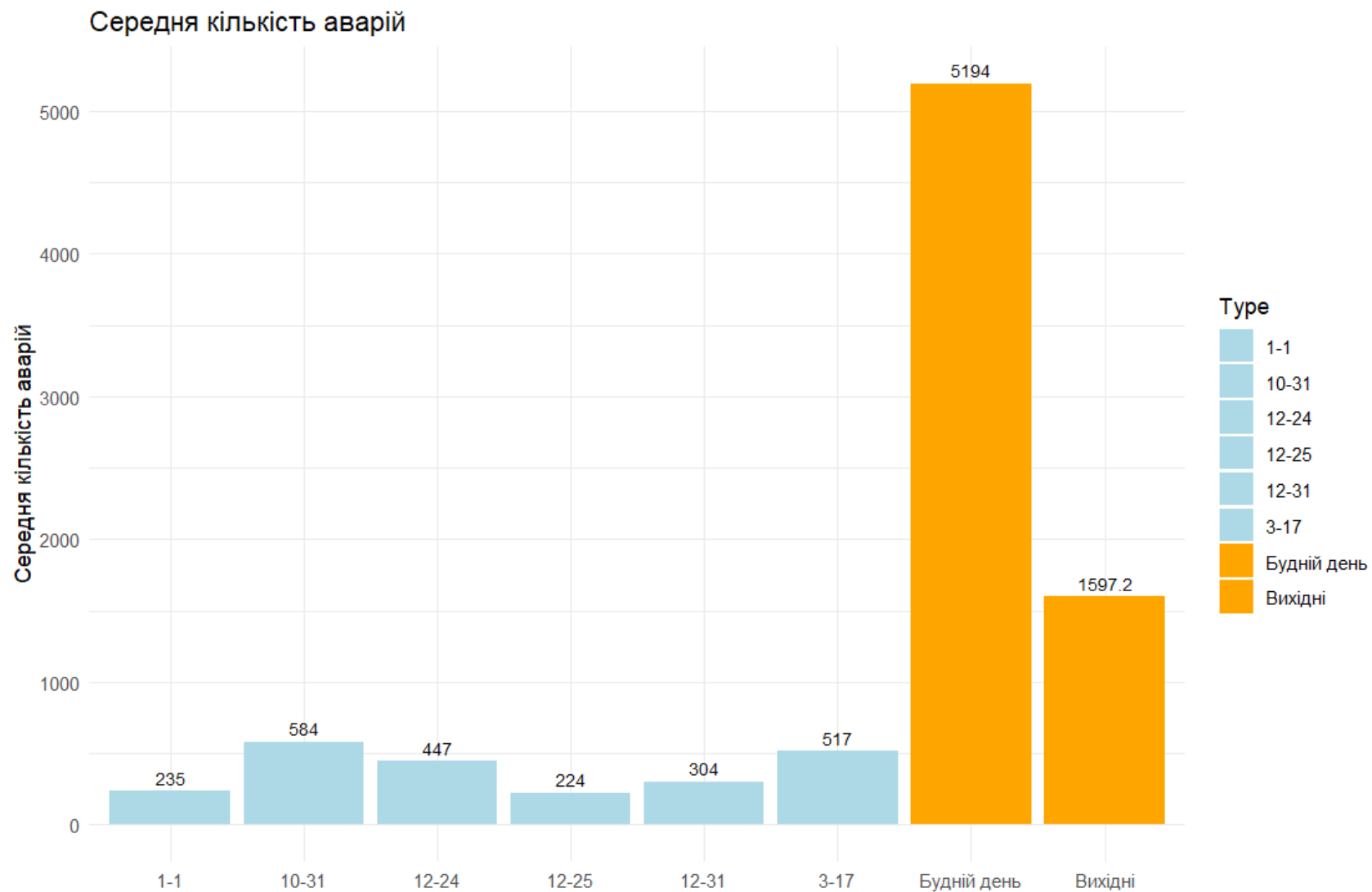


Топ-10 вікових категорій водіїв
за кількістю смертельних аварій



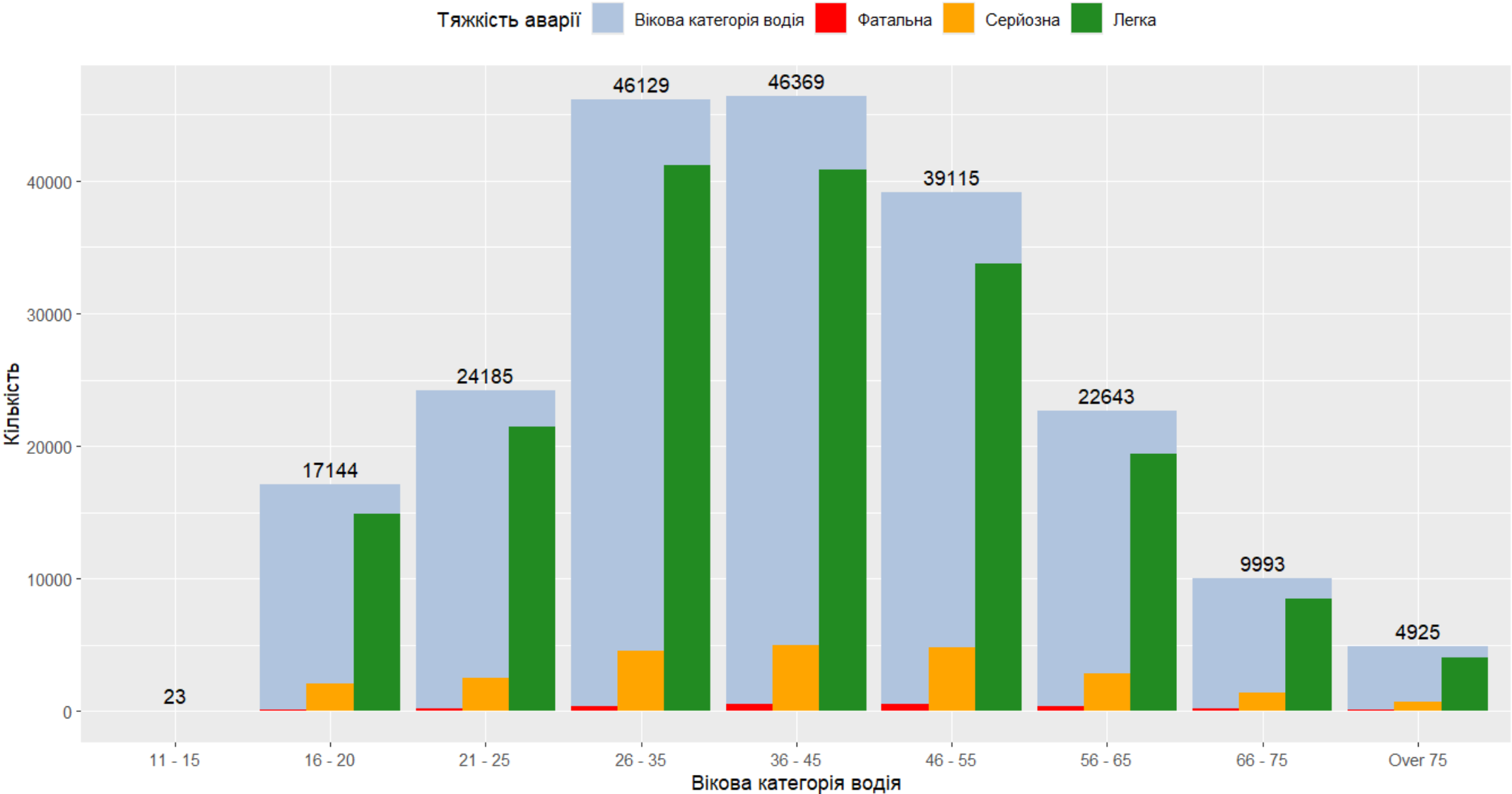
3. Як впливає час/день/рік на кількість/серйозність аварій?



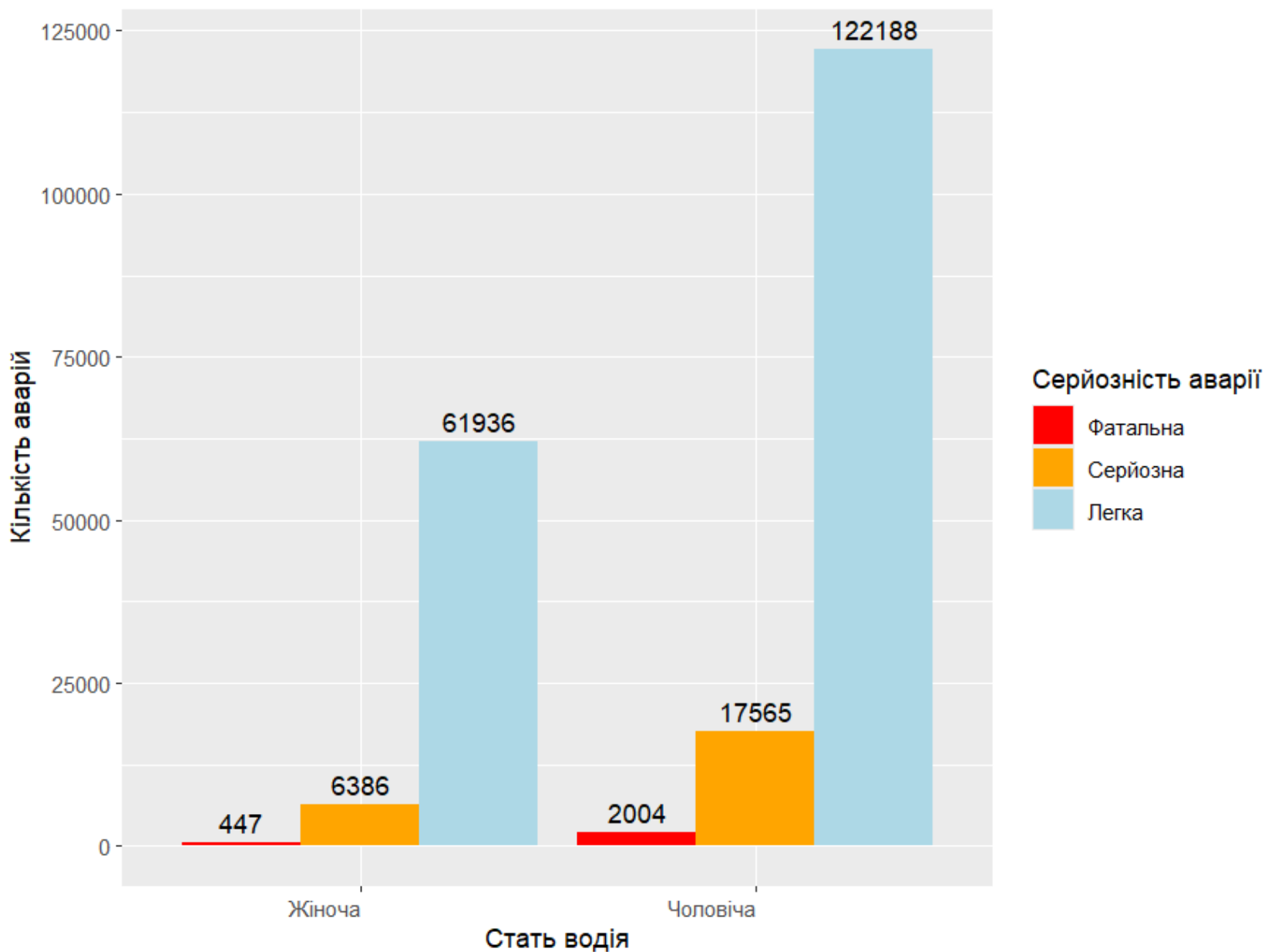


У вихідні дні середня кількість ДТП суттєво менша, ніж у будні дні. А у свята ця кількість ще менша, при цьому найменше ДТП трапляється у різдвяні свята, а найбільша - напередодні Нового року.

4. Вплив фізичних особливостей водія на поведінку на дорозі?

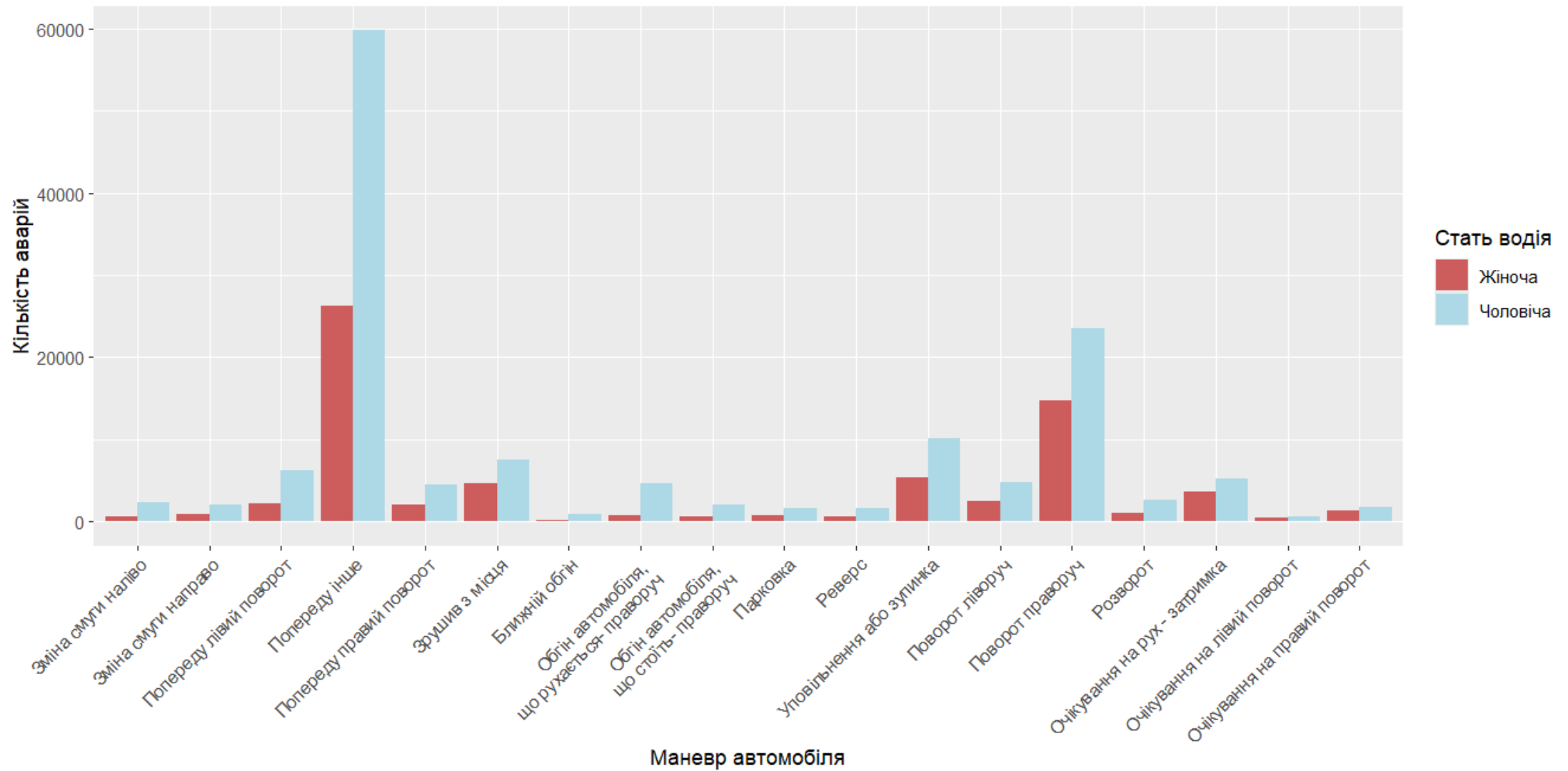


Кількість ДТП за статтю водія та тяжкістю

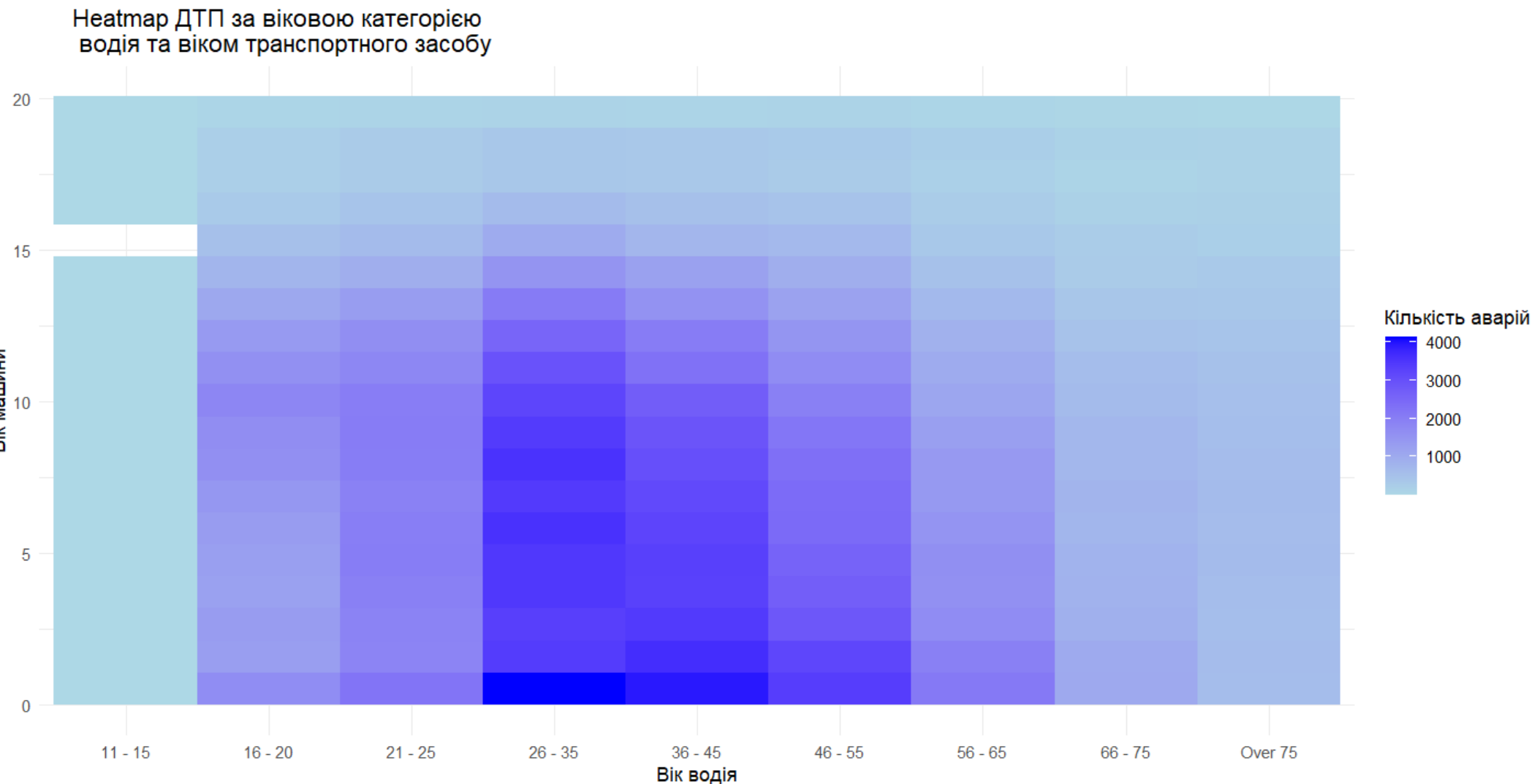


З офіційних ресурсів ми можемо дізнатись про те, що кількість виданих ліцензій на водіння за весь час для чоловіків 25956385, а для жінок 23438355. Отже, чоловіки складають приблизно 52.49% від всіх виданих ліцензій на водіння, а жінки - приблизно 47.51%. Тим не менше, чоловіки частіше потрапляли в аварії, причому частка смертельних ДТП у них майже в 4.5 рази більша, ніж у жінок, а частка серйозних ДТП майже в 3 рази більша.

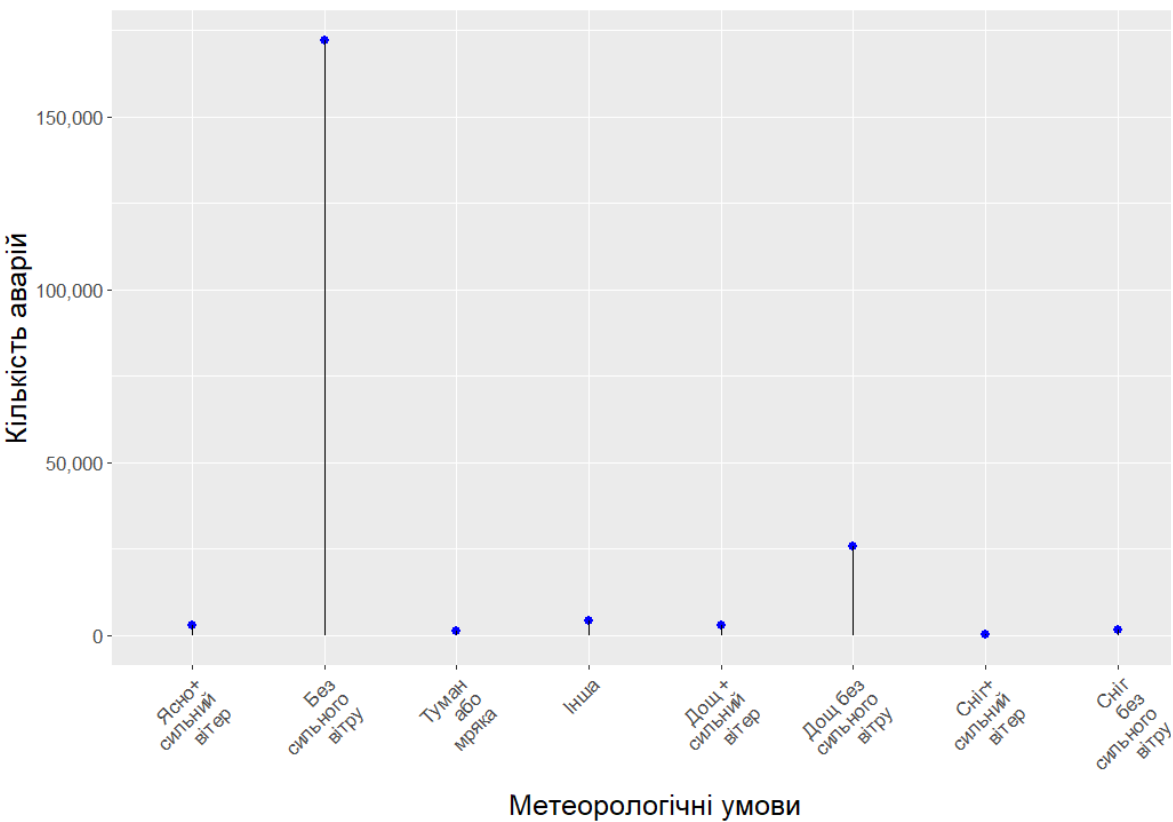
Розподіл маневрів автомобіля за статтю водія



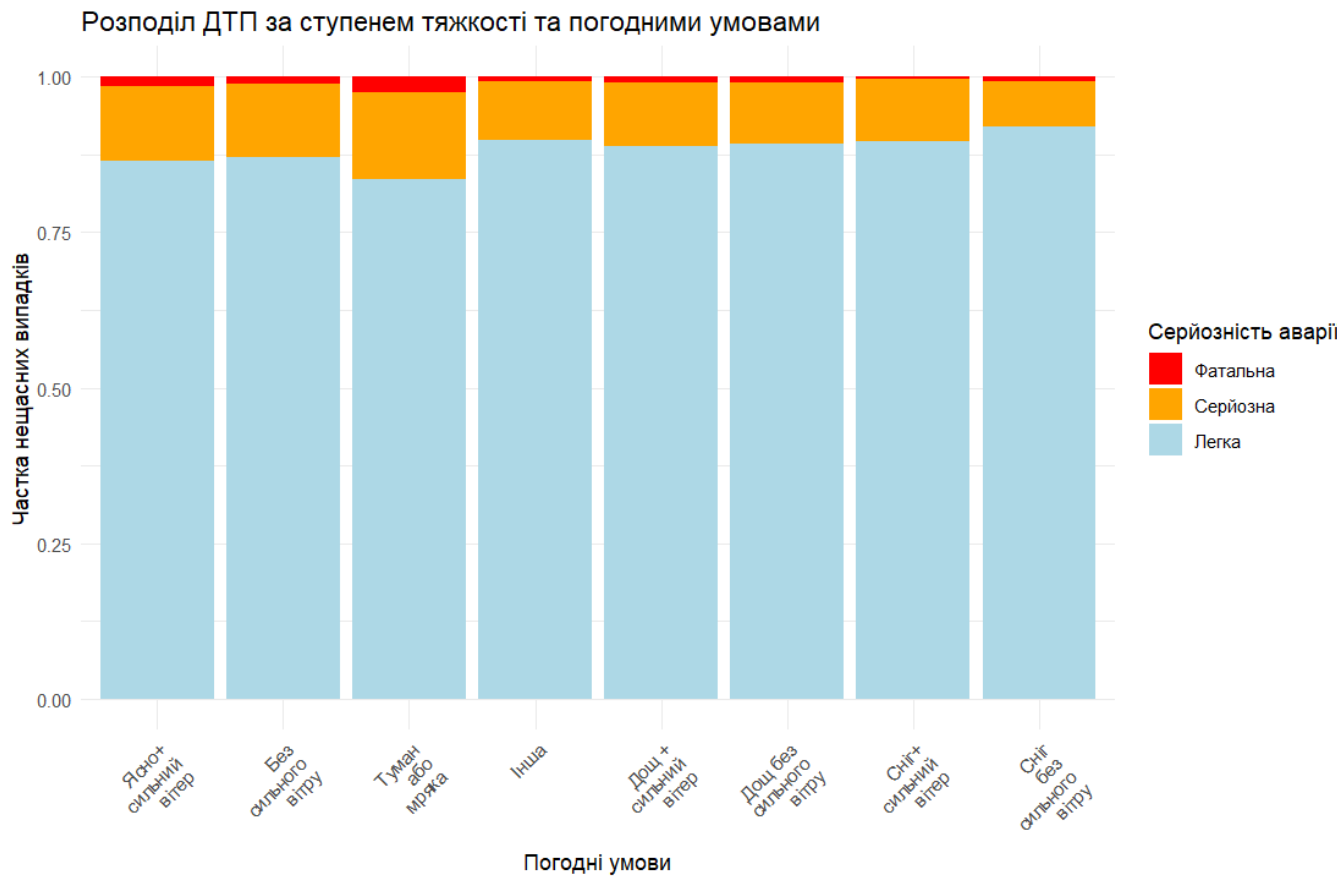
Найбільша кількість аварій людей віком від 26-45 років та віком машини до 2 років



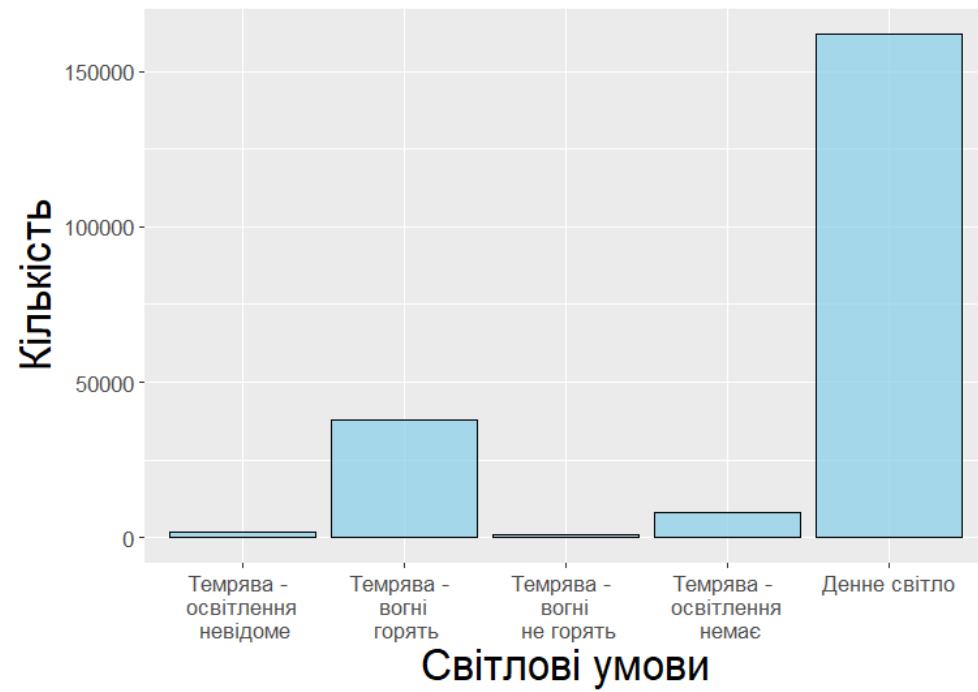
5. Як фактори на які не може впливати водій впливають на кількість /серйозність аварій?



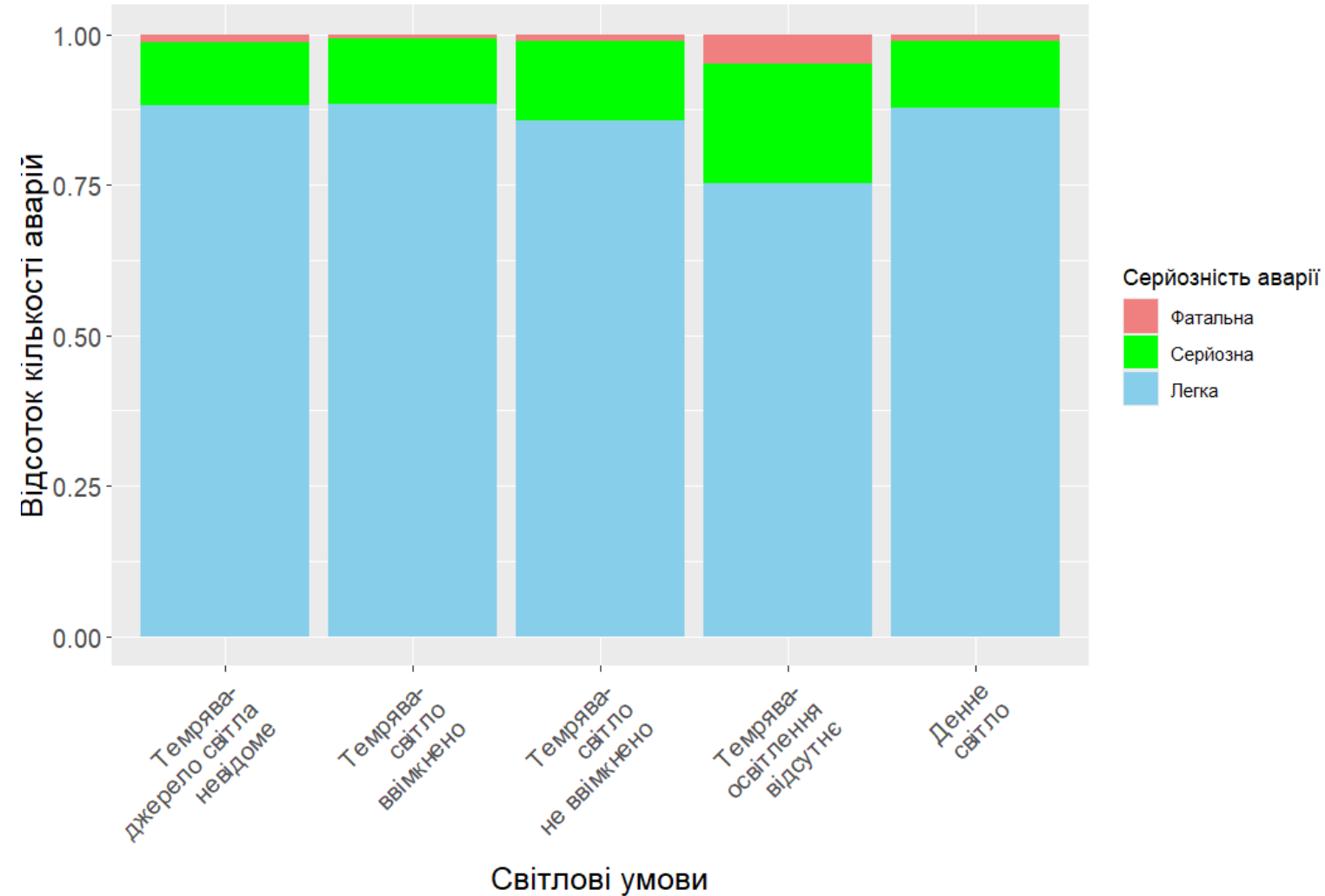
Найбільше аварій трапляється коли немає сильного вітру або дош без сильного вітру



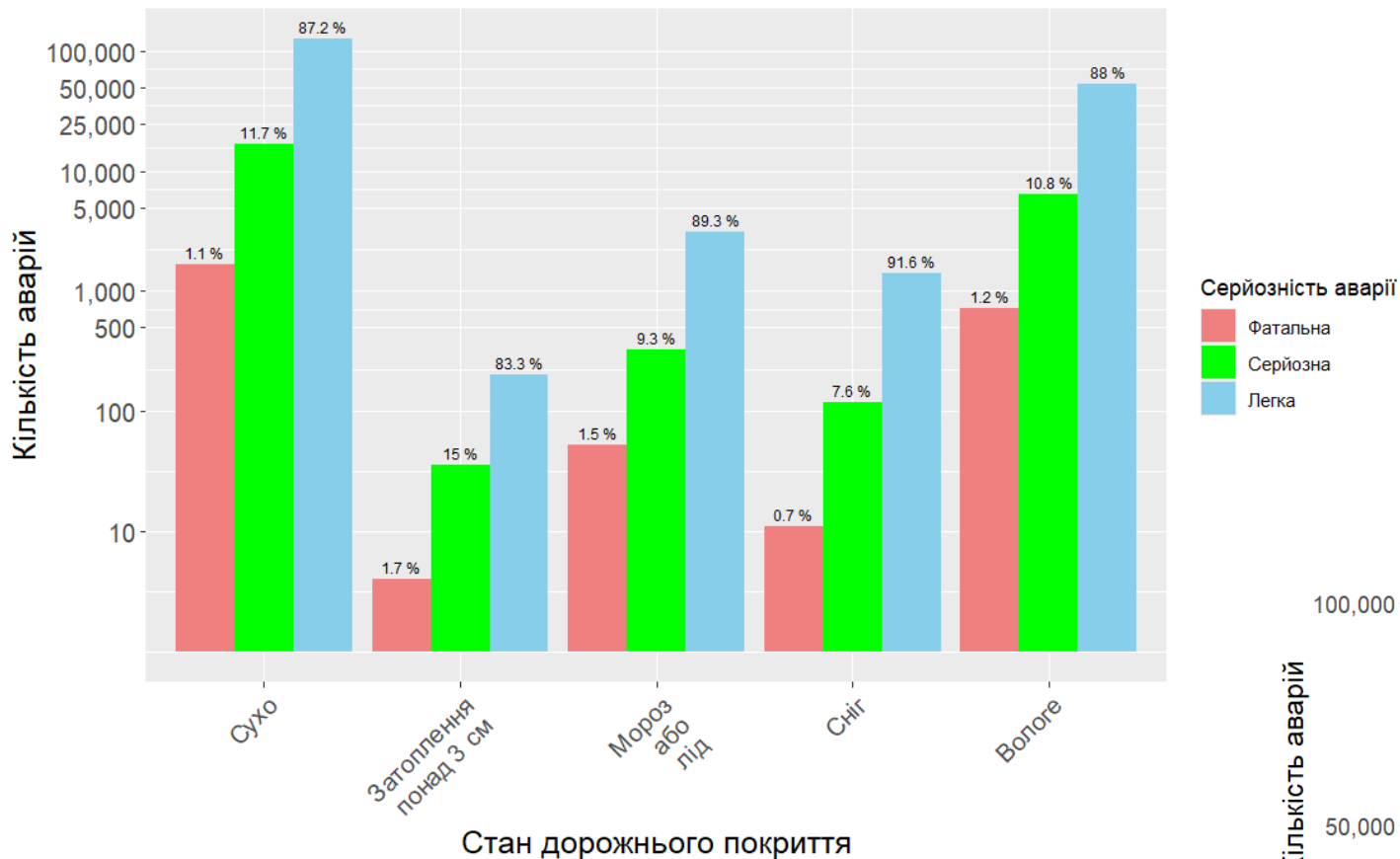
При тумані або мряці найбільше фатальних аварій



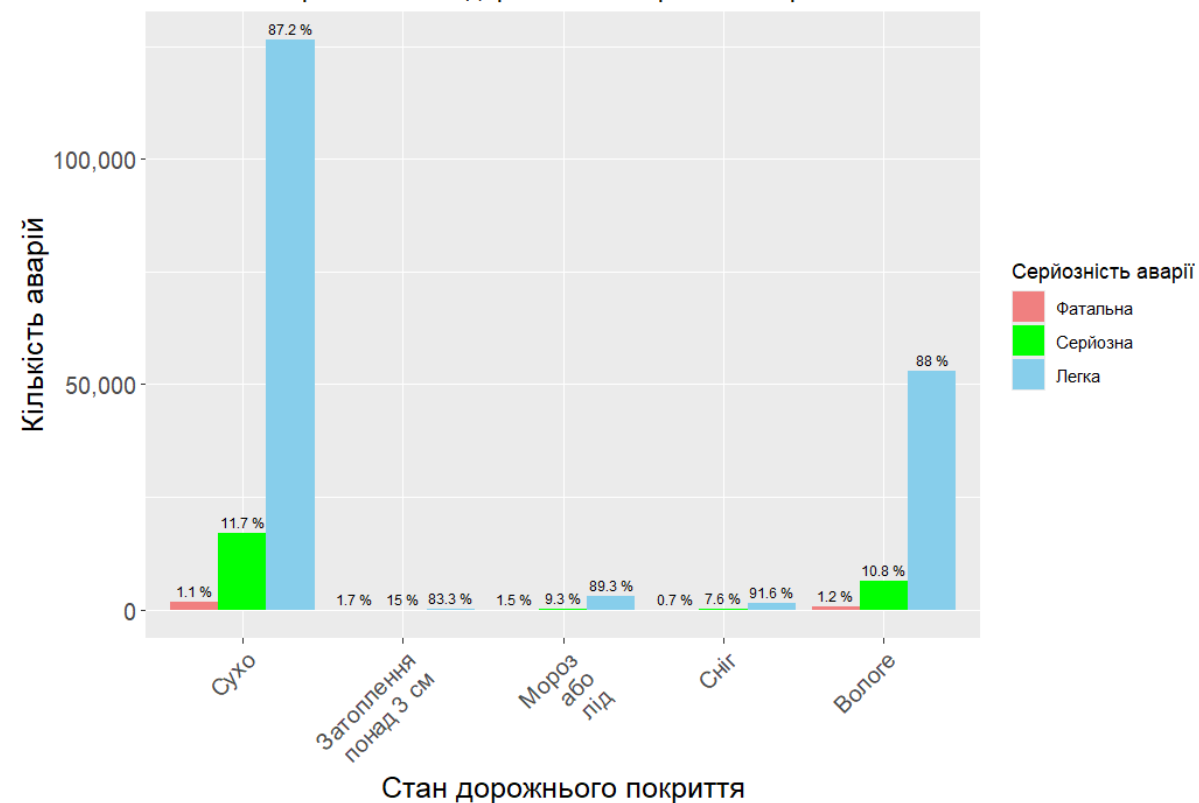
Близько 170 000 аварій трапляються при денному світлі



Кількість аварій за станом дорожнього покриття та серйозністю(логарифмізований)



Кількість аварій за станом дорожнього покриття та серйозністю

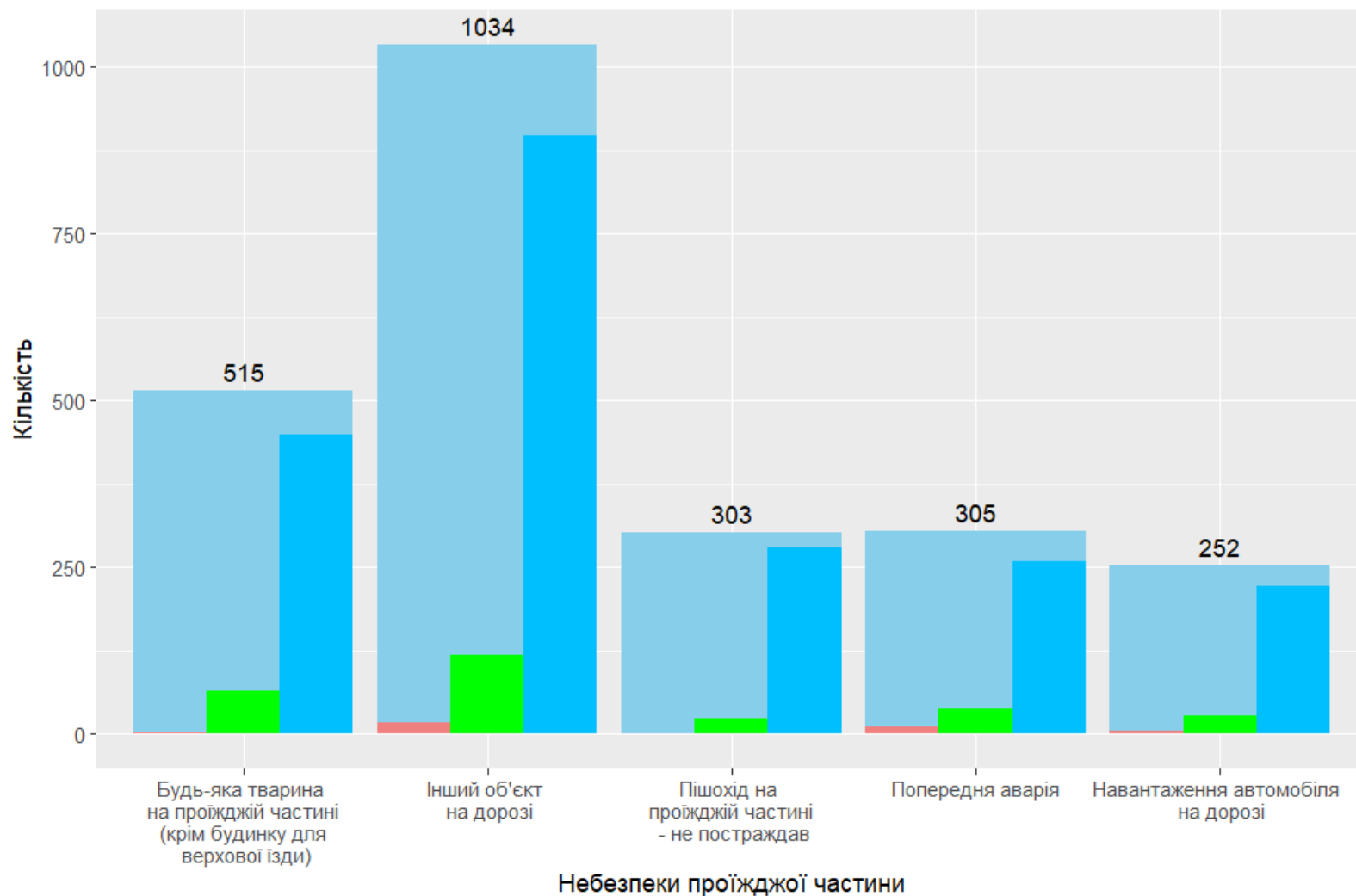


Серйозність аварії

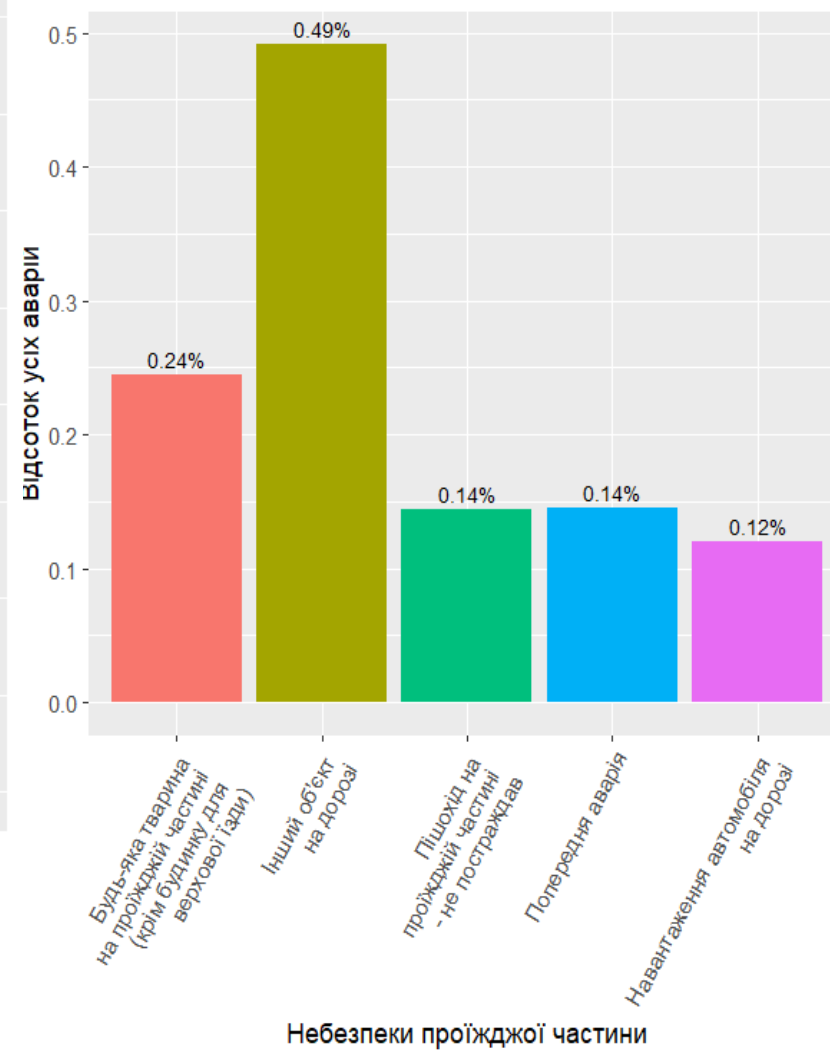
Фатальна

Серйозна

Легка



Розподіл небезпек на проїжджій частині



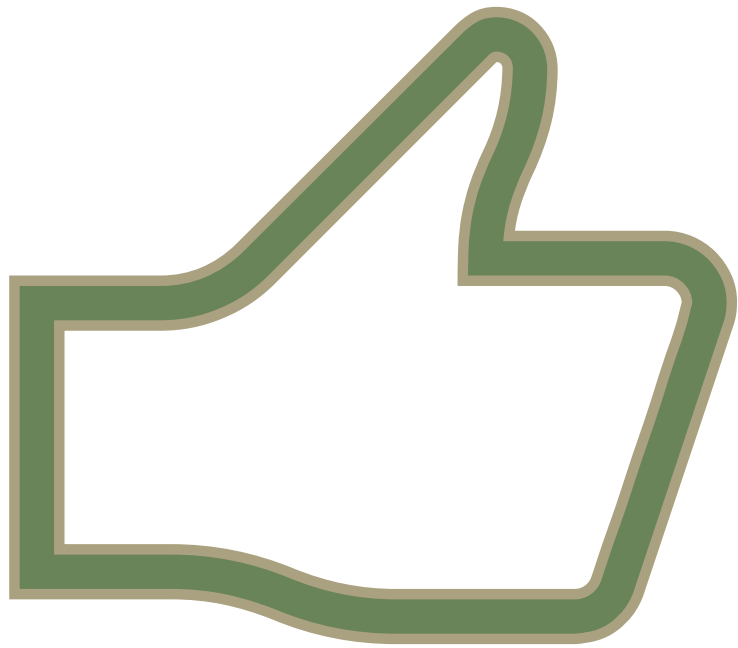
ВИСНОВКИ

У датасеті присутні записи ДТП, в яких Age_of_Vehicle дорівнює 63 або 86, однак прямої залежності між віком транспортного засобу і ймовірністю потрапляння на ньому в ДТП не було виявлено.

Не можна впевнено казати про фактори, що спричинили ДТП, оскільки вони не є незалежними. Фатальна аварія могла статись вночі, без освітлення в дощ на поганій дорозі, але казати, що всі три фактори рівнозначно вплинули на ДТП не можна. За деякими виключеннями, всі фактори розглядалися окремо.

Загалом, найбільший вплив на тяжкість і ймовірність потрапляння ДТП є:

- місцевість (в сільській місцевості кількість фатальних ДТП є значно вищою, ніж в міській, і навпаки, кількість легких ДТП в місті перевищує такі в сільській місцевості).
- погодні умови (ДТП під час туману мають найбільше співвідношення фатальних ДТП до легких і серйозних, далі йдуть ДТП під час дощу і снігу)
- світлові умови (ДТП під час темряви, де транспортний засіб не увімкнув фари складають найбільшу частку фатальних і серйозних)



Дякуємо за увагу!