

lab_1

Gil Shiloh and Dean Teslter

30 3 2022

R Markdown

שאלה 1

גרף 1

גרף 1 מתאר לנו את מספר התאונות במטרופולין של חיפה בחודשים בשנים 2013 - 2019. בנוסף קיים לנו קו ממוצע שמראה לנו את מספר התאונות הממוצע בחודש בתקופת השנים.

א

הגרף מנסה לענות לנו על השאלה: האם יש מגמה וקשר כלשהו בין חודשים ותקופות עוקבות במטרופולין של חיפה.

ב

במבט על הגרף קשה להגיד שהגרף עונה על השאלה שכן אין מגמה ברורה בגרף וישנם קפיצות רבות בין החודשים. בנוסף נראה שברוב התקופות מספר התאונות קרוב לממוצע.

ג

הגרף משאיר הרבה שאלות פתוחות שמצריכות להיכנס לתוך הנתונים על מנת למצוא את התשובות שלהן.. לדוגמה הגרף רק מציג את מספר התאונות בזמן שבתקופה זאת האוכלוסיה גדלה, כלומר לא מופיע יחס התאונות לאזרחים. האם יש סיבה לכך שבשנה הראשונה גרף ישנם קפיצות גדולות. האם יש חודשים שבהם היו תאונות גדולות עם הרוגים או מספר גבוה של פצועים קשה לעומת חודשים עם תאונות קלות יותר.

ד

לדעתינו היה אפשר להציג בגרף את מספר התאונות שבהם היו מעורבים פצועים קשה והרוגים ובנוסף את שאר התאונות על מנת להראות את חומרת התאונות. בנוסף היה אפשר להוסיף קו רגרסיה על מנת שנוכל לראות ביתר קלות האם קיימת מגמה של מספר התאונות.

גרף 2

גרף זה מציג לנו את מספר התאונות במטרופולין חיפה שבהן יש פציעות בהשוואה בין מבוגרים לילדים ומציג לנו את מגזרי הערים במטרופולין. בנוסף יש לנו מעין קו רגרסיה שמציג לנו את הקשר בין תאונות של מבוגרים וצעירים.

א

הגרף מנסה לענות לנו על השאלה האם יש קשר בין מספר התאונות בהקשר של פציעות של צעירים ומבוגרים במטרופולין חיפה ובנוסף האם יש קשר בין מגזר לבין מספר התאונות.

ב

נראה שהגרף עונה בצורה טובה על השאלה. נראה שיש קשר בין מספר התאונות של מבוגרים וצעירים. בנוסף בצורה ברורה יותר ניתן לראות שישנה מגמה ברורה שבא בערבים מהמגזר הערבי מספר התאונות בקרב צעירים הוא גבוה מאוד, לעומת ערים יהודיות ומעורבות שבהן ישנו קשר די קבוע בין מספר התאונות שבהם מעורבים צעירים לעומת מבוגרים.

ג

השאלה שעולה לנו מדוע דווקא ביישובים ערביים יש מעורבות גדולה יותר של צעירים. ניתן לשער שאחוז הילדים ביישובים אלו הוא גדול יותר מאשר ביישובים יהודיים ולכן עקב כך שאחוזם באוכלוסיה הוא גדול יותר כך גם מספר התאונות בקרבם הוא גדול יותר. אך לא בהכרח ויתכן שישנם סיבות נוספות ואחרות.

ד

לכן על מנת לשפר את הגרף היינו מציעים להוסיף את אחוז הצעירים והמבוגרים בקרב היישובים על מנת לקבל השוואה טובה יותר, במקום רק לקבל את מספר התאונות כנתון.

שאלה 2

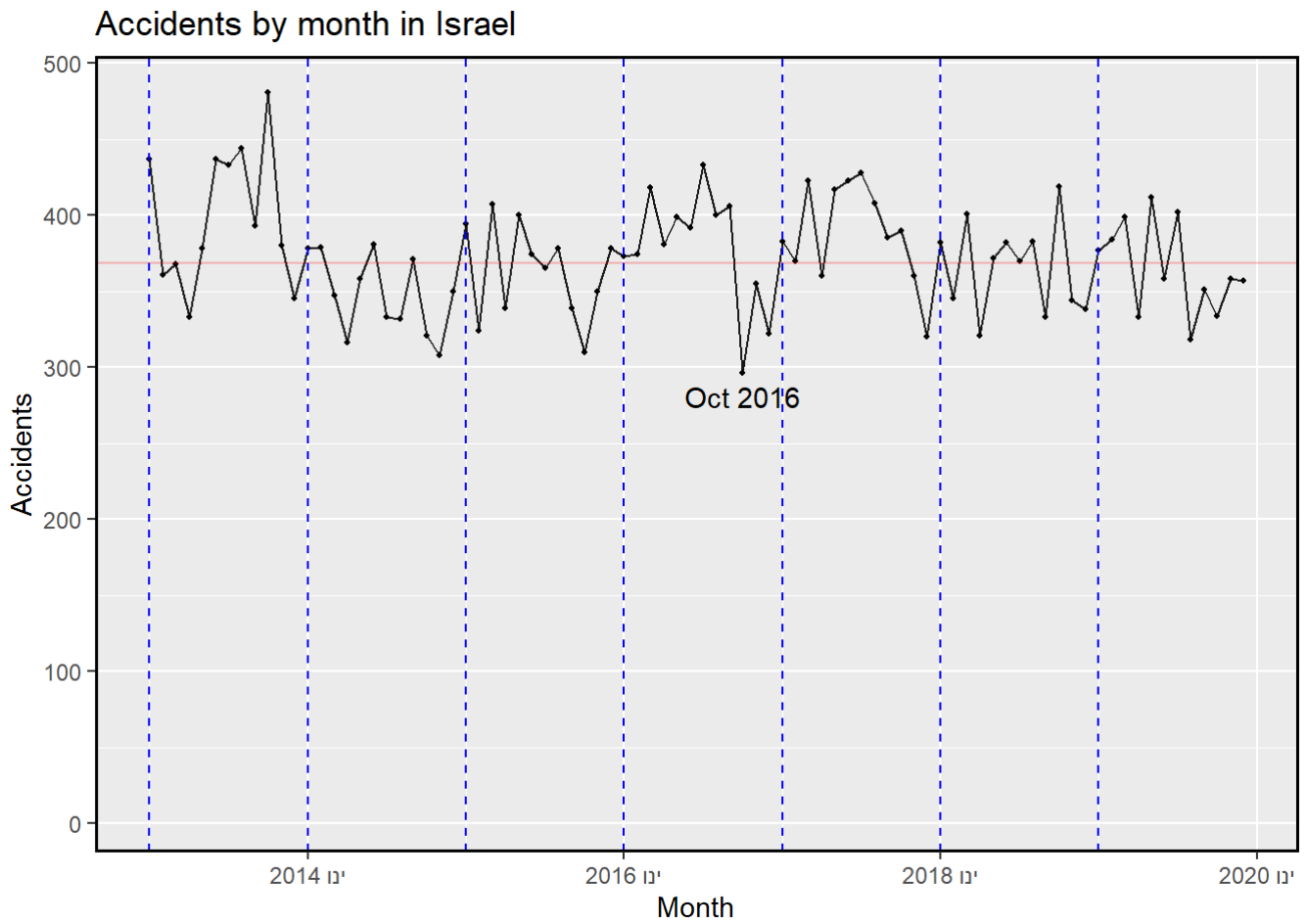
גרף - 1

```
## `summarise()` has grouped output by 'Year'. You can override using the `.groups`  
## argument.
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

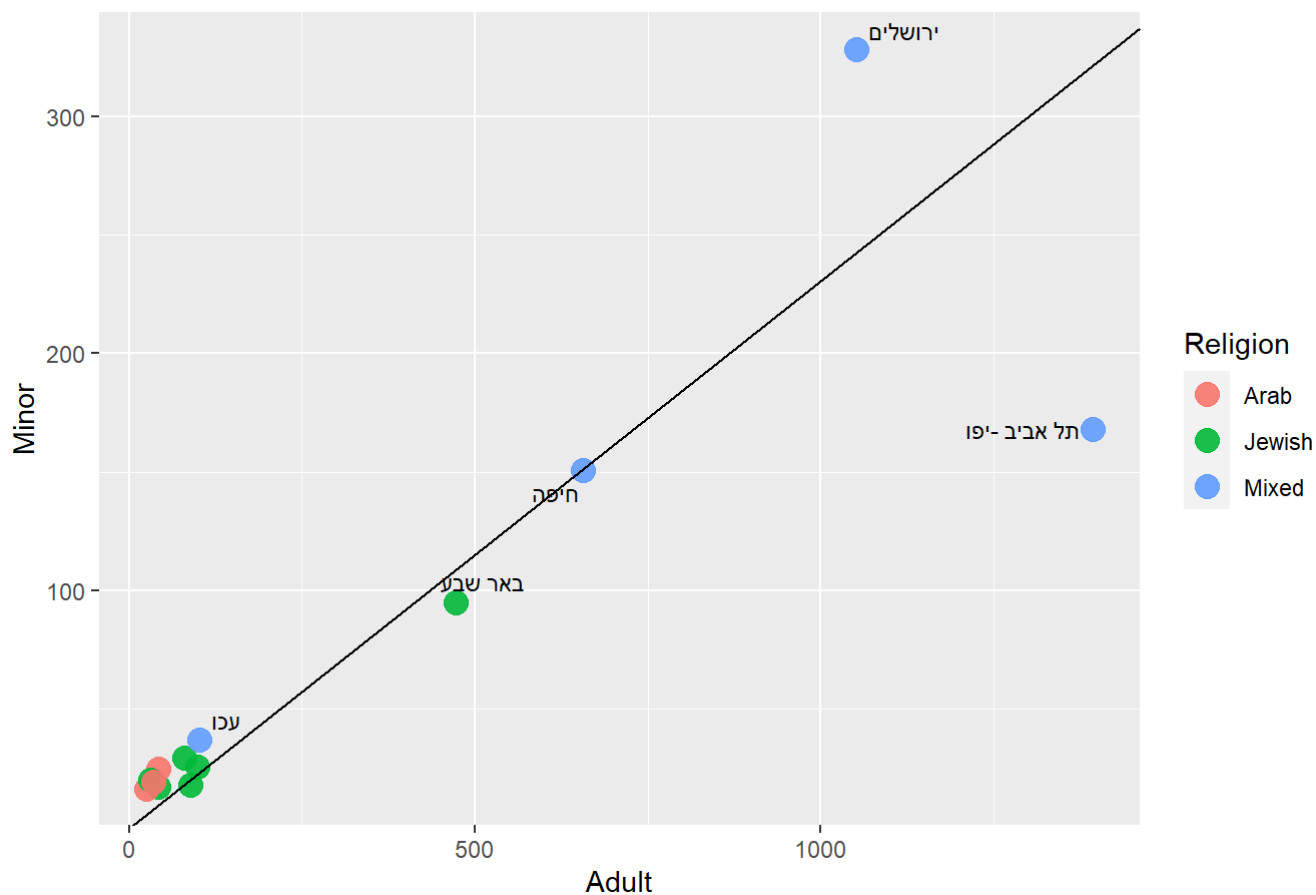
```
## Warning: Removed 1 rows containing missing values (geom_text).
```



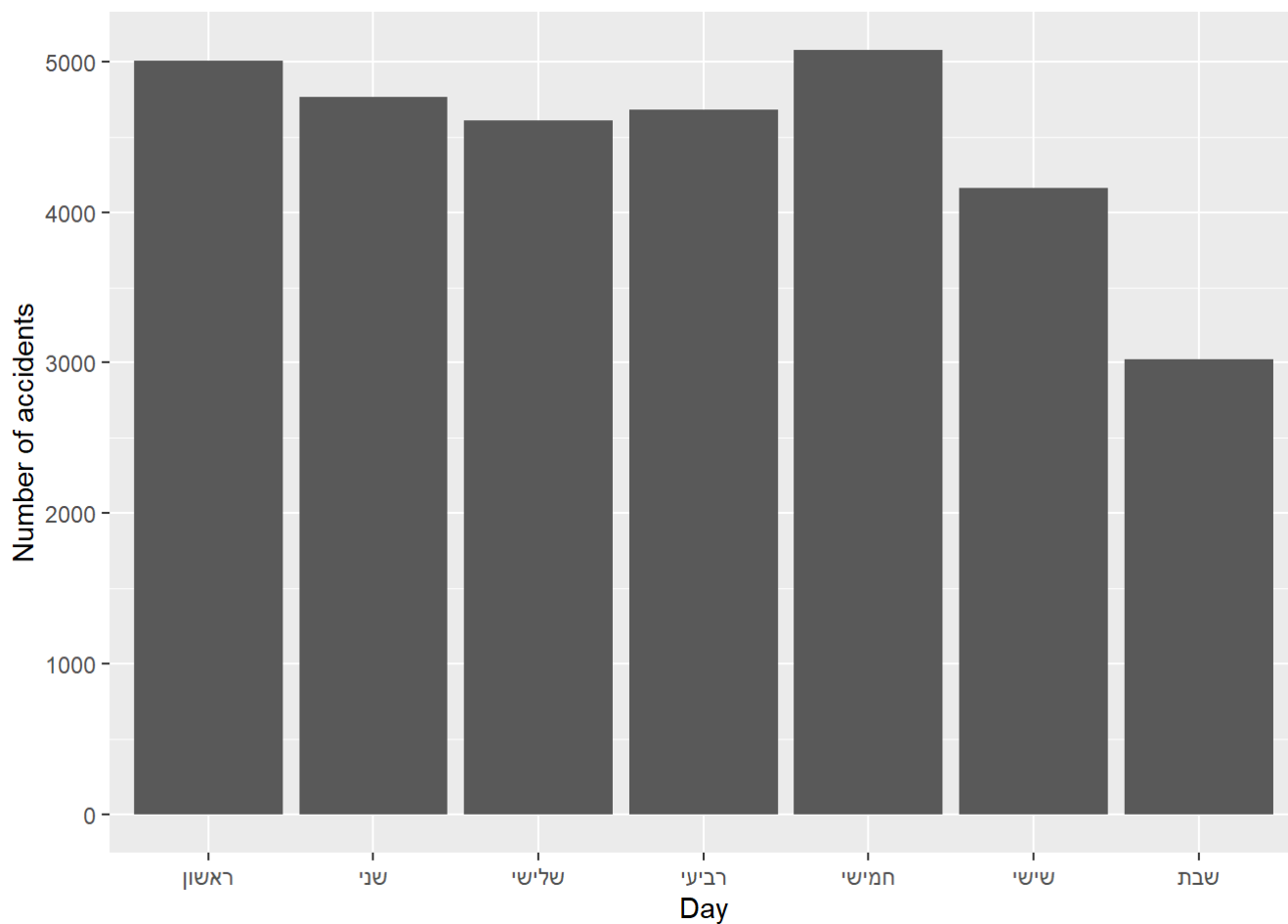
גרף - 2

```
## Warning: ggrepel: 8 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

Total accident-related injuries in town by age (2013-2019), yearly rate

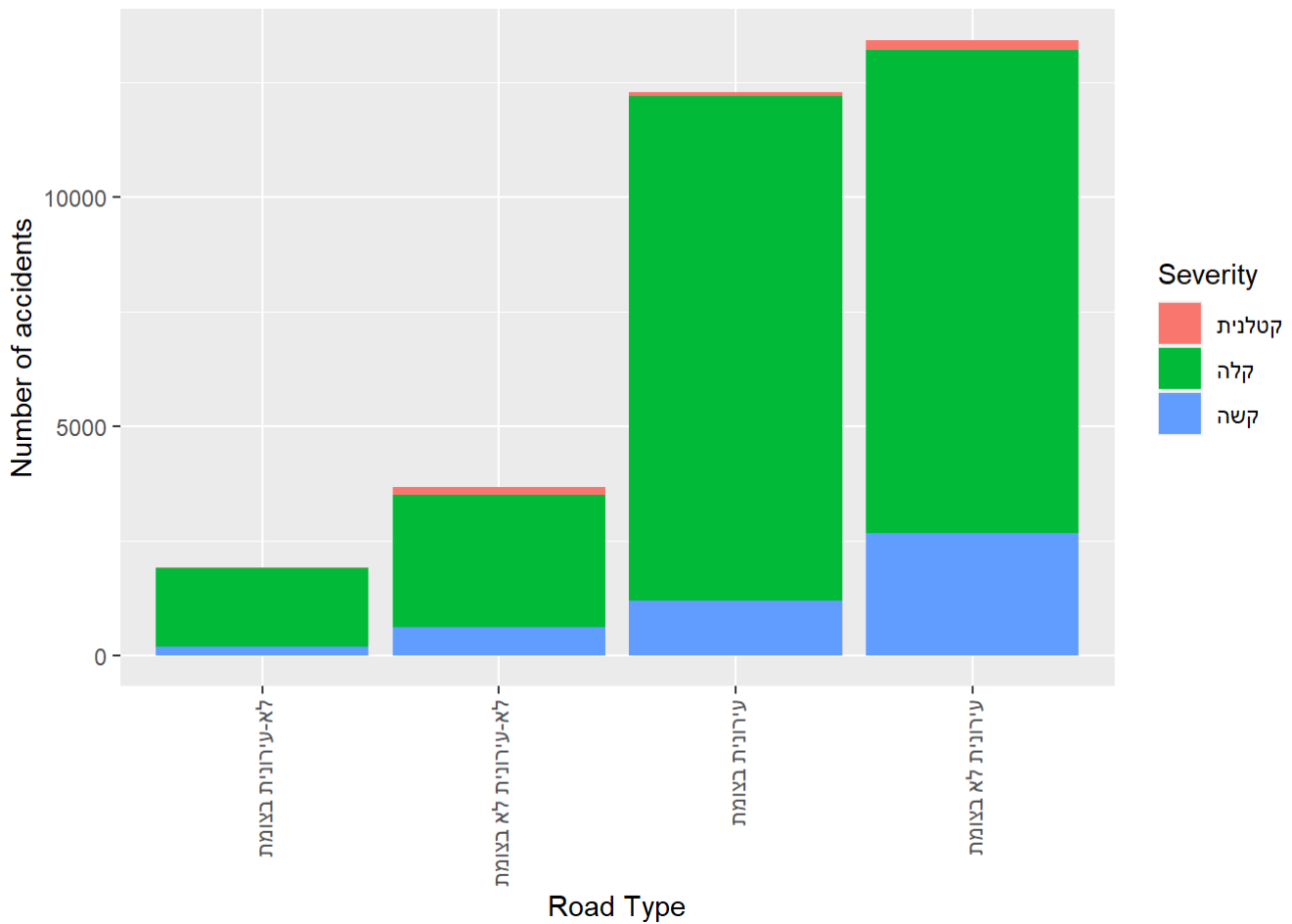


שאלה 3



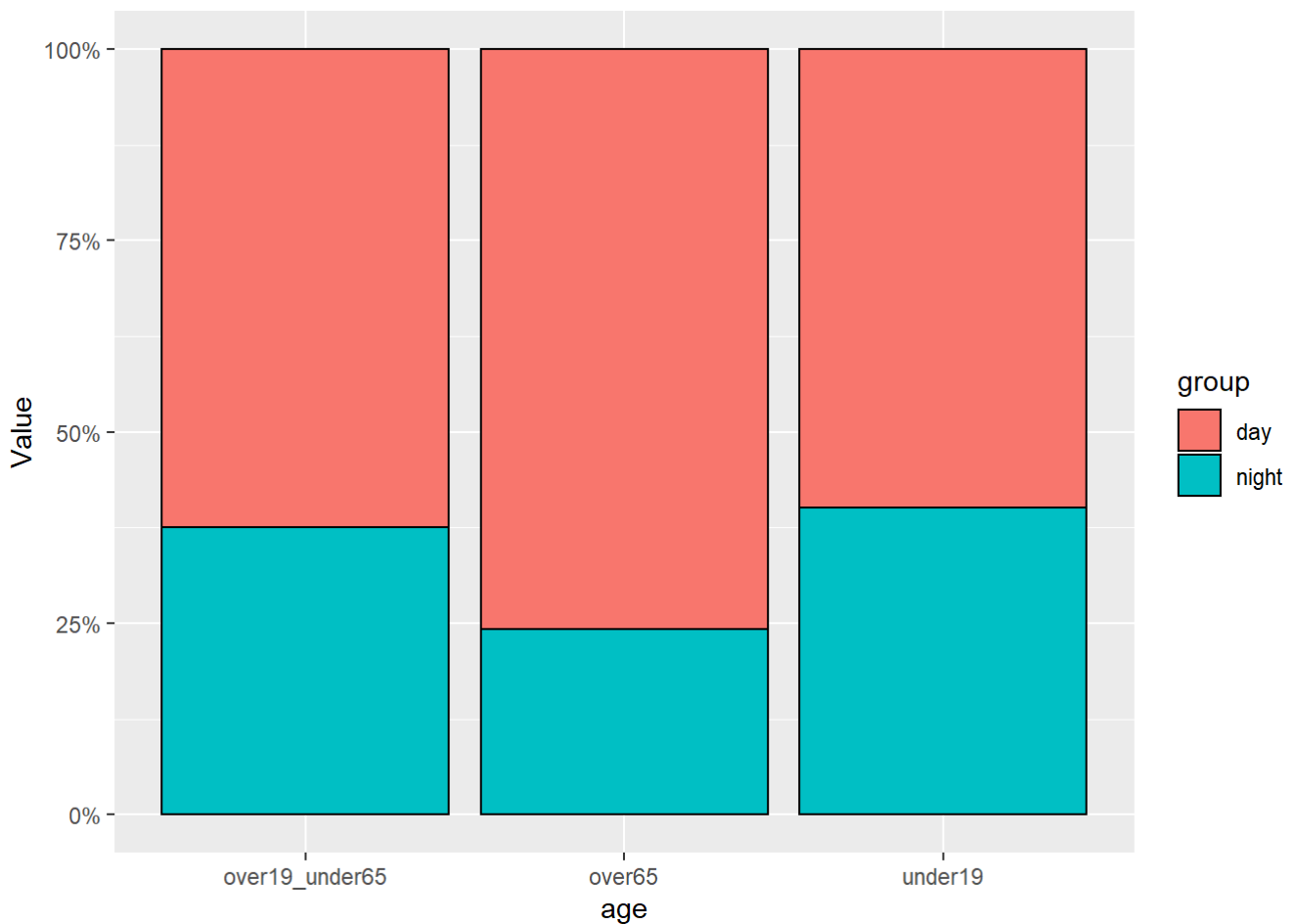
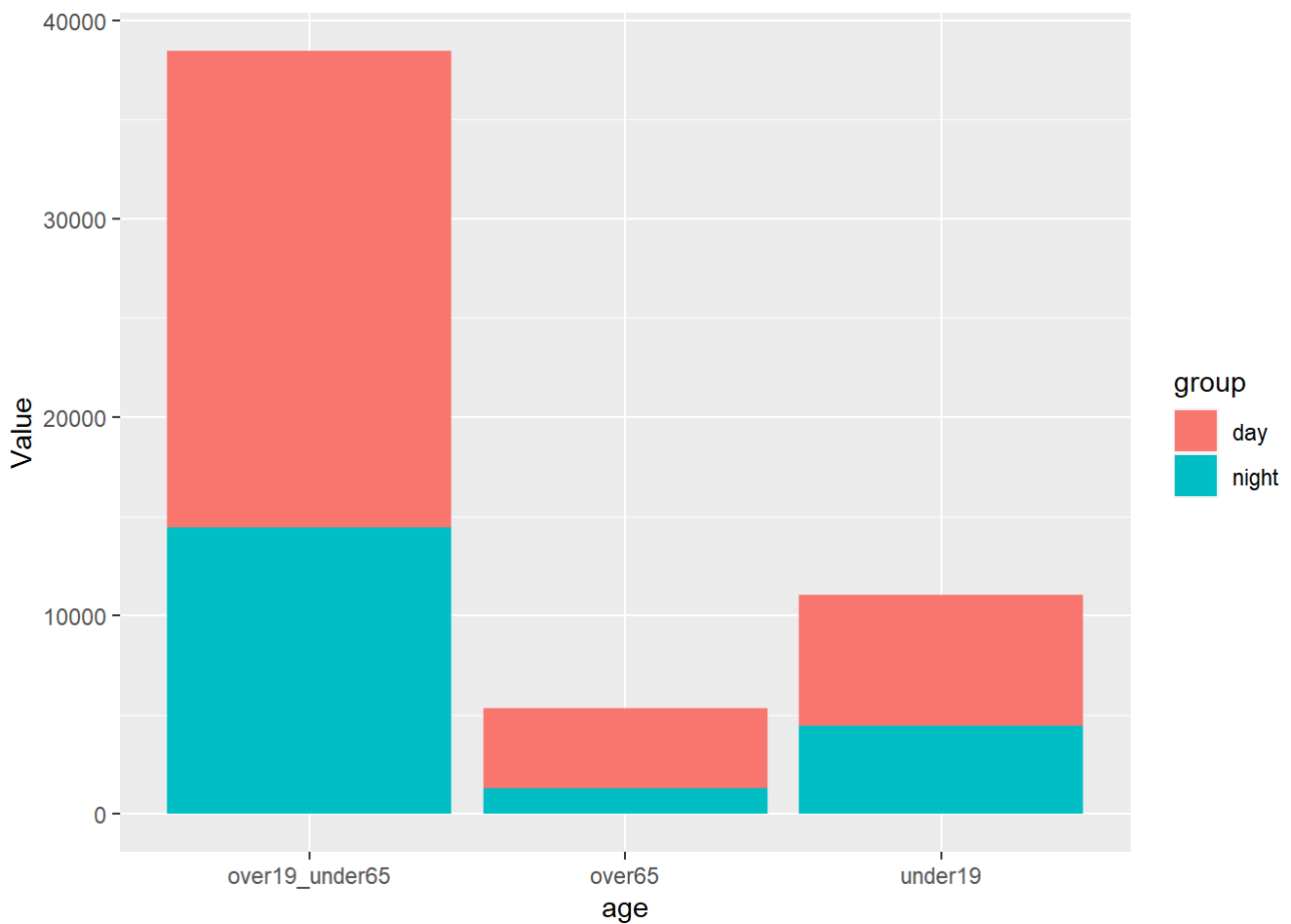
מהגרף ניתן לראות שלאורך השבוע מספר התאונות דומה חוץ מימי שישי ושבת. אנו מניחים שהבדל זה נובע מכמות פחותה של מכוניות שנמצאות על הכביש עקב אנשים דתיים אשר בימי שישי מרגע כניסת השבת הם אינם נוסעים במכונית וכך יש פחות מכוניות על הכביש וכדומה בימי שבת לאורך היום.

```
## `summarise()` has grouped output by 'Road_Type'. You can override using the
## `.groups` argument.
```



בניתוח זה בחרנו לבדוק את חומרת התאונות ואת כמות התאונות אל מול סוג הדרך שבה קרתה התאונה. ניתן לראות כי רוב התאונות קורות בדרכים עירוניות וכן גם יותר תאונות קשות בהתאם. בנוסף ניתן להבחין בהבדל בין צמתיים/לא צמתיים - ניתן לראות שקורה יותר תאונות לא בצמתיים והתאונות יותר חמורות בדרכים אשר אינם צמתיים.

```
##   group      age Value
## 1  day    under19  6614
## 2 night    under19  4438
## 3  day over19_under65 24012
## 4 night over19_under65 14441
## 5  day    over65   4027
## 6 night    over65  1287
```

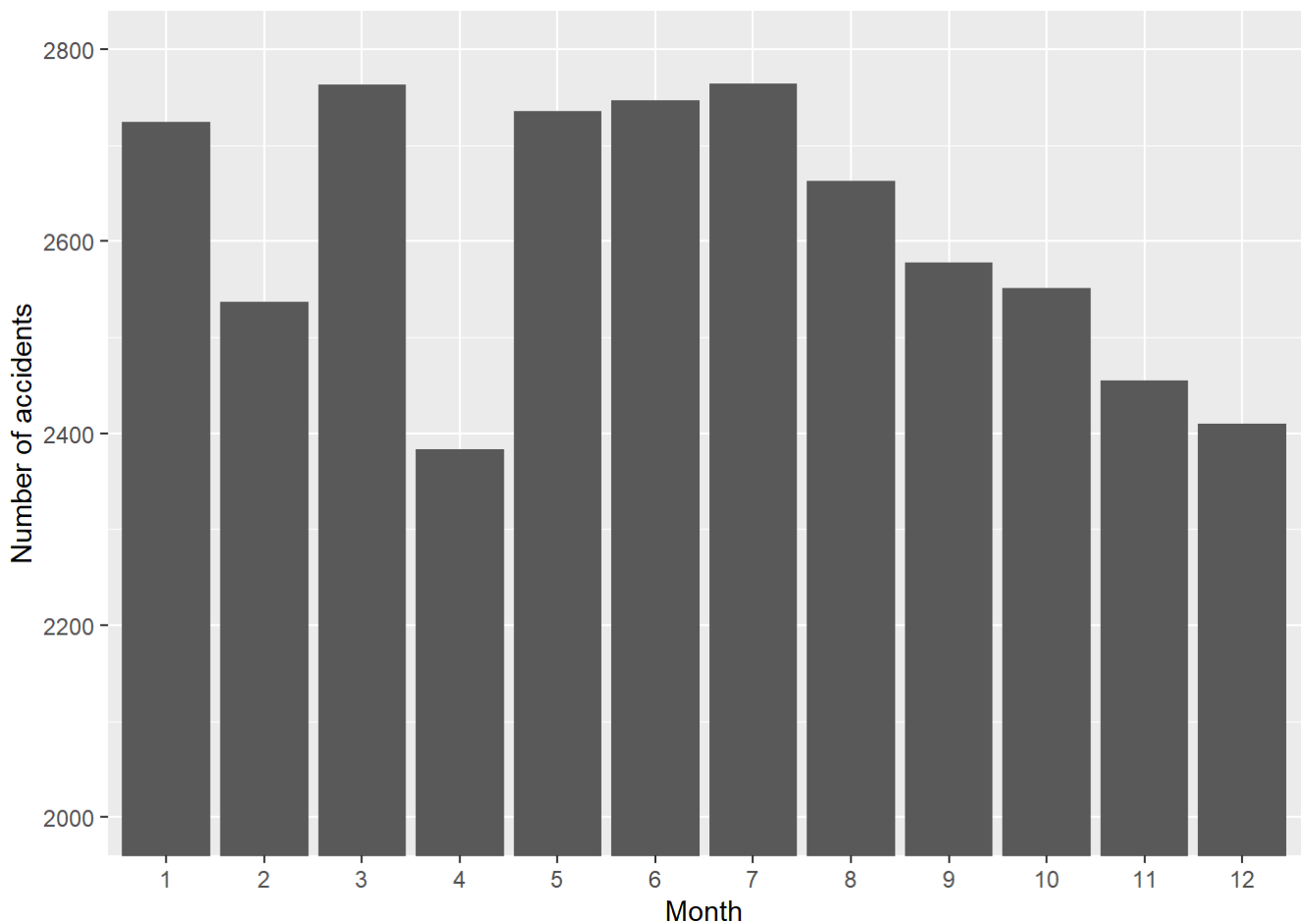


מהגרפים הללו ניתן לראות כמה דברים שאף מפתיעים. ציפינו לראות שבקבוצת הגיל של 19 - 65 יהיו הכי הרבה תאונות. אך אנחנו יכולים לראות מהגרפים את מספר ואחוז התאונות של קבוצות הגיל ביום לעומת הלילה. נתייחס לכל קבוצה בנפרד. בקבוצת הגיל של 65 פלוס אין הפתעה שלקבוצה זאת הכי מעט תאונות שכן הם נמצאים מעט יחסית על הכבישים ובדרכים. בנוסף ניתן

לראות שבקרבת התאונות של גיל 65 פלוס רוב התאונות קורות ביום, ושוב נתון זה אינו מפתיע כי לרוב מבוגרים אינם נמצאים בשעות הלילה על הכביש וברחובות. בקבוצת הגיל של מתחת ל-19, שהיא קבוצת הגיל במקום השני מבחינת מספר התאונות הופתענו מאוד לגלות את אחוז התאונות שבהם מעורבים קבוצת הגיל הזאת בשעות הלילה ותאונות הלילה מהוות כ-40 אחוז מסך התאונות של קבוצת גיל זאת. עקב כך שלרוב חברי קבוצת גיל זאת אין רישיון נהיגה, לא היינו מצפים לראות אותם מעורבים בתאונות רבות בלילה. חברי קבוצת גיל זאת מעורבים בתאונות לילה באחוז גבוה יותר מאשר כל שאר הקבוצות, נתון מפתיע. לבסוף יש לנו את קבוצת הגילאים של 19-65 ובאופן לא מפתיע לקבוצה זאת מספר התאונות הגדול ביותר שכן הם מונים את מספר האנשים הגדול ביותר ובנוסף קבוצת גילאים זאת נמצאת הכי הרבה על הכבישים. בקבוצת גילאים זאת אחוז התאונות בלילה הוא כמעט 40 אחוז ונתון זה אינו מפתיע שכן ישנן נסיעות רבות בשעות הלילה.

שאלה 4

1



accidents by month



מהגרפים לא ניתן להסיק שהשערת האפס נכונה, כלומר שהתפלגות הנתונים אינה מגיעה מהתפלגות רנדומלית - יוניפורמית. הגרף הראשון מייצג את התפלגות הנתונים, ניתן להבחין בקו מגמה יורד בחודשים האחרונים שמייצגים את חודשי החורף ויש תלות בכך ולכן אנו מסיקים כי מהגרפים לא ניתן לקבוע שהנתונים הגיעו מהתפלגות יוניפורמית.

שאלה 5

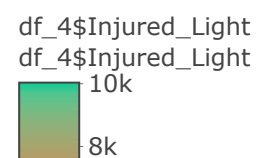
```
## Importance of first k=3 (out of 6) components:
##
##          PC1      PC2      PC3
## Standard deviation  4197.3836 303.88250 57.90656
## Proportion of Variance  0.9944  0.00521 0.00019
## Cumulative Proportion  0.9944  0.99964 0.99983
```

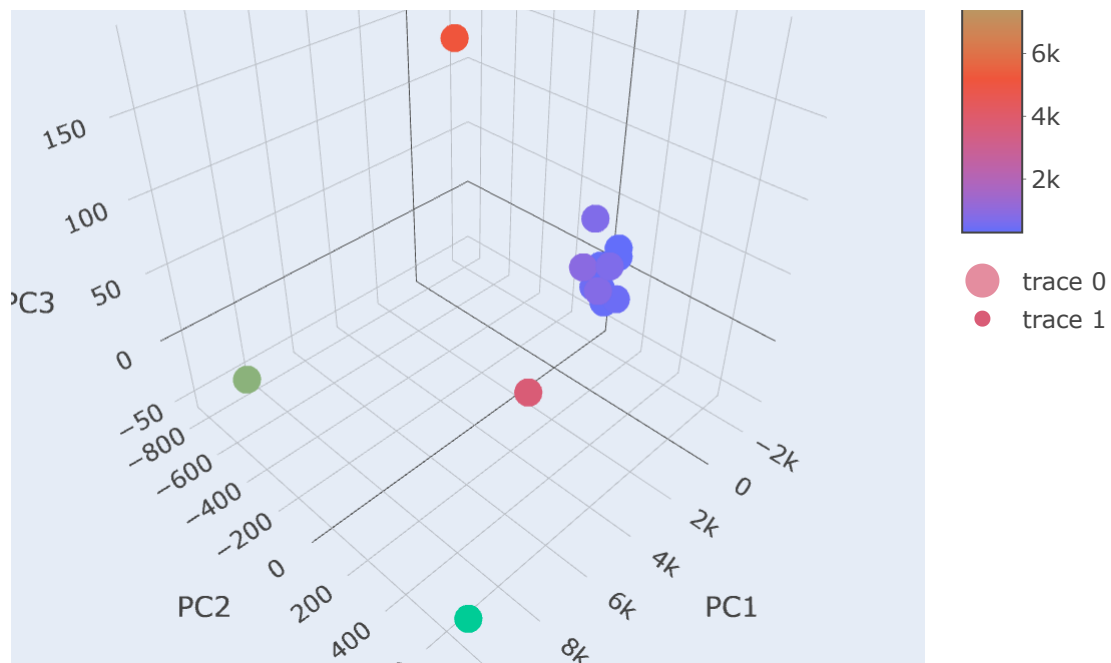
מהטבלה אנו רואים שהפיסי הראשון הוא הרלוונטי ביותר לנו והוא מכיל את החלק הגדול ביותר שלנו - 0.994.

```
## No trace type specified:
## Based on info supplied, a 'scatter3d' trace seems appropriate.
## Read more about this trace type -> https://plotly.com/r/reference/#scatter3d
```

```
## No scatter3d mode specified:
## Setting the mode to markers
## Read more about this attribute -> https://plotly.com/r/reference/#scatter-mode
```

3D PCA plot





מנתוני הגרף התלת מימדי אנחנו יכולים לראות שיש לנו קבוצה גדולה של ערים שהם מקובצים. אם נסתכל על ערים אלה נוכל לראות שאלו יישובים וערים קטנות יחסית. לעומת זאת יש לנו את תל אביב, באר שבע, חיפה וירושלים שערים אלו נמצאים רחוק מההתקבצות של השאר. ניתן לשאר שהערים הגדולות מתקבצות רחוק בגלל שלהן יש את מספר הפצועים וההרוגים הגדול ביותר. ואלו הם שמשיעים על הפיסיאי