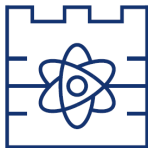


Analiza danych Ubera w New York
kwiecień - wrzesień 2014



K.S. R.A.

Politechnika krakowska im. Tadeusza Kościuszki
Wydział fizyki

Używane biblioteki

Obórka danych

Analiza danych

Heatmaps

Używane biblioteki



Pierwszym etapem pracy był import niezbędnych pakietów potrzebnych do analizy:

- ▶ ggplot2 - wizualizacji danych
- ▶ ggthemes - dodatek do ggplot
- ▶ lubridate - nasz zestaw danych obejmuje różne ramy czasowe. Aby zrozumieć nasze dane w osobnych kategoriach czasowych, skorzystamy z pakietu lubridate
- ▶ dplyr - pakiet ten jest językiem od manipulacji danych w R
- ▶ tidyr - pomaga upořadkować dane
- ▶ DT - pozwala łączyć się z biblioteka JavaScript o nazwie - Datatables
- ▶ scales - za pomocą skal graficznych możemy automatycznie mapować dane do prawidłowych skal za pomocą dobrze rozmieszczonych osi i legend



Obórka danych



Kolejnym krokiem było wczytanie danych i podział je na odpowiednie ramki danych.

Dane podzieliśmy na obiekty zależne od czynników czasowych, takich jak dzień, miesiąc, rok, dzień tygodnia itp.



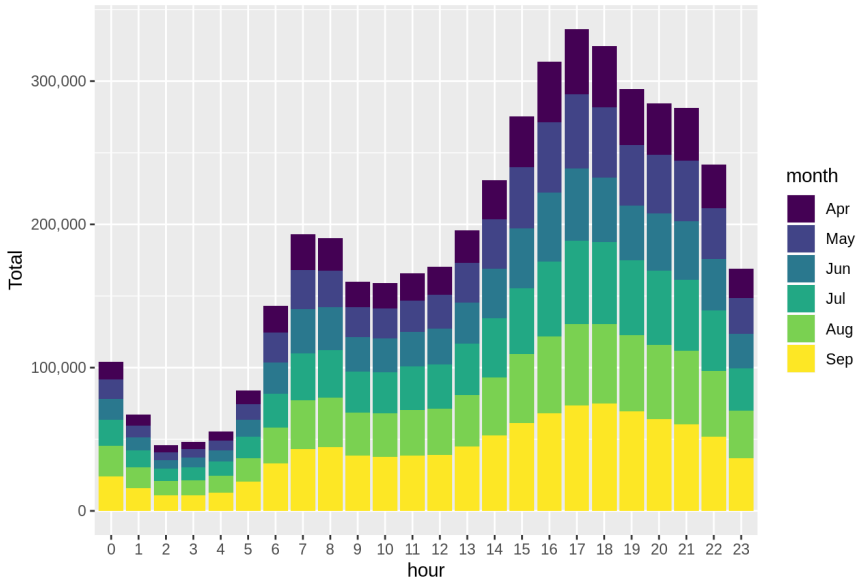
Analiza danych



Dokonaliśmy obliczeń i pierwszy wykres jaki został narysowany przedstawia w jakich godzinach pasażerowie odbyli najwięcej przejazdów w każdym z miesięcy.



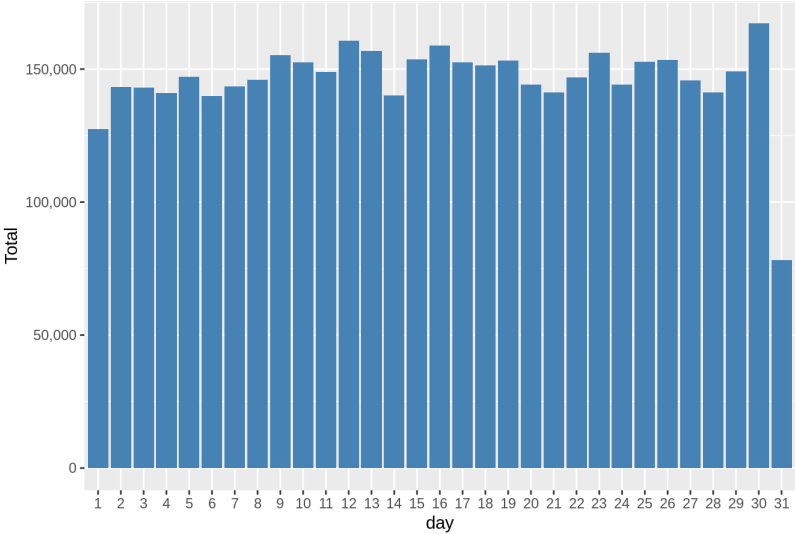
Trips by Hour and Month



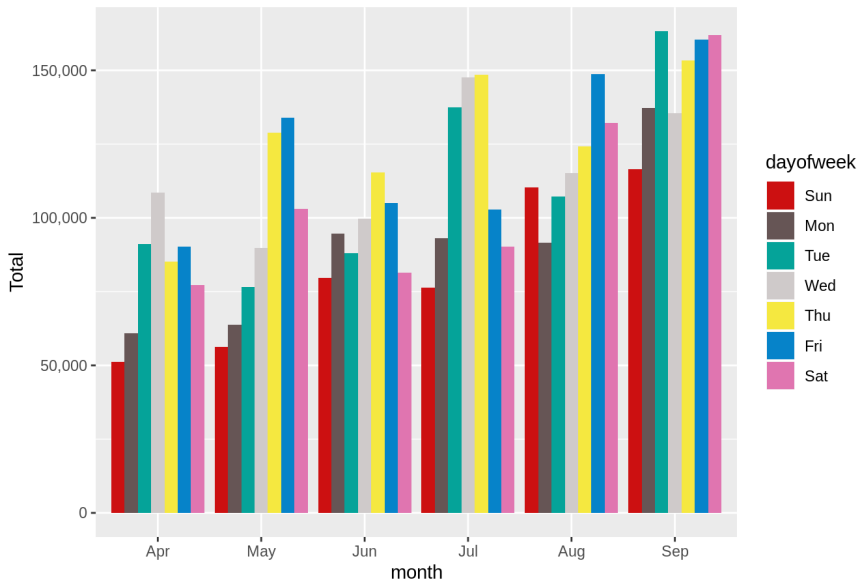
W kolejnej części projektu podzieliliśmy dane na podstawie
każdego dnia miesiąca.



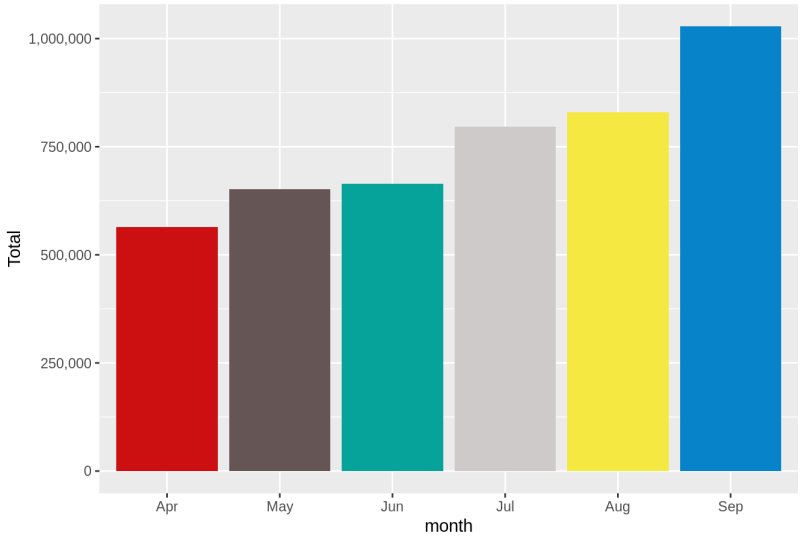
Trips Every Day



Trips by Day and Month



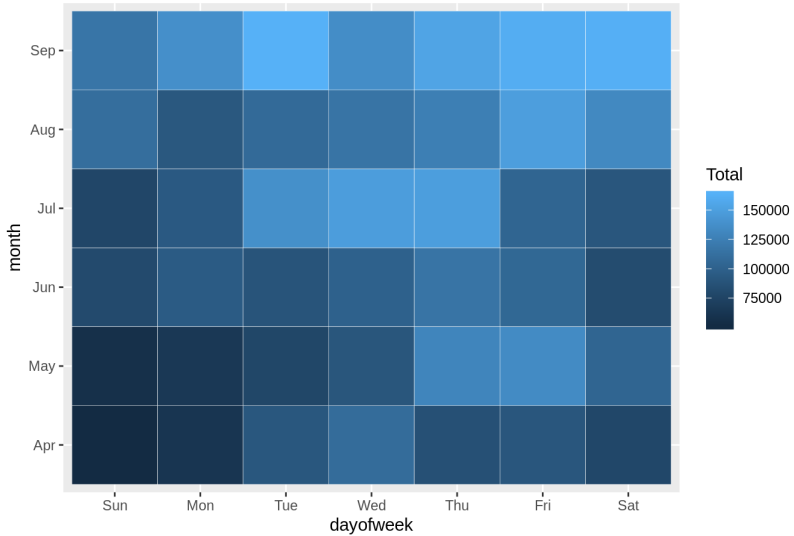
Trips by Month



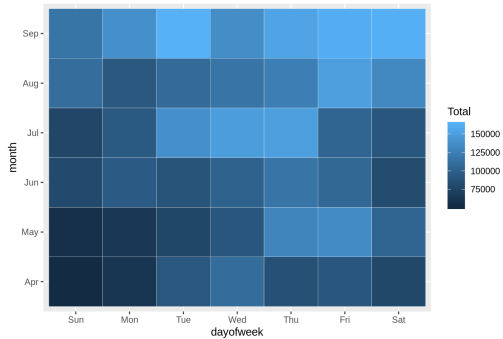
Heatmaps

- ▶ dla godziny i dnia.
- ▶ dla miesiąca i dnia.
- ▶ dla miesiąca i dnia tygodnia.

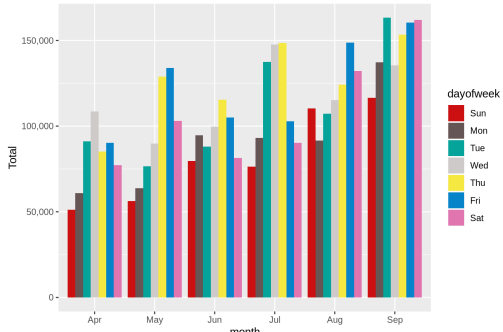
Heat Map by Month and Day of Week



Heat Map by Month and Day of Week



Trips by Day and Month



W prezentacji zobaczyliśmy wizualizacje danych w języku R. Wykorzystaliśmy pakiety takie jak ggplot2, które pozwoliły nam wykreślić różne rodzaje wizualizacji, które dotyczyły kilku ram czasowych w roku. Dzięki temu możemy stwierdzić, w jakim czasie klienci Ubera przemieszczali się.



Dziękujemy
