TheLorry

Jaeseok Jeong
Department of Statistics, SKKU

September 2, 2018

1.배송 출발지역 시각화

쿠알라 룸푸르(Kuala lumpur) 지역에 밀집

다른 지역도 뭉쳐있는 부분은 아마 대도시일듯 (말레이시아 지역을 잘몰라서 우선..)

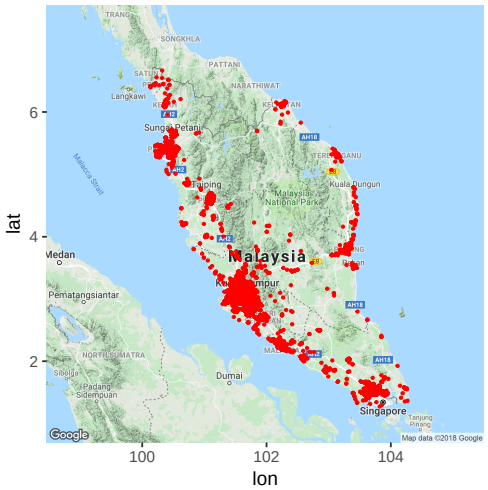
도시 내 이동이 많은지/ 도시 외부로 이동하는지 지역별로 살펴봐야할듯

```
pickup_location <- location_1 %>%
  dplyr::filter(type == "pickup")

map <- get_map(location = 'Malaysia', zoom = 7)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=Malaysia&zoom=7&size=640x640&scale=2&m
```

```
## Source : https://maps.googleapis.com/maps/api/geocode/json?address=Malaysia
```

```
pickup_map <- ggmap(map) +
  geom_point(aes(x = lng, y = lat),
             data = pickup_location,
             size = 0.5, col = "red")

pickup_map
```



2. 도시 외 이동 개별 id 예시

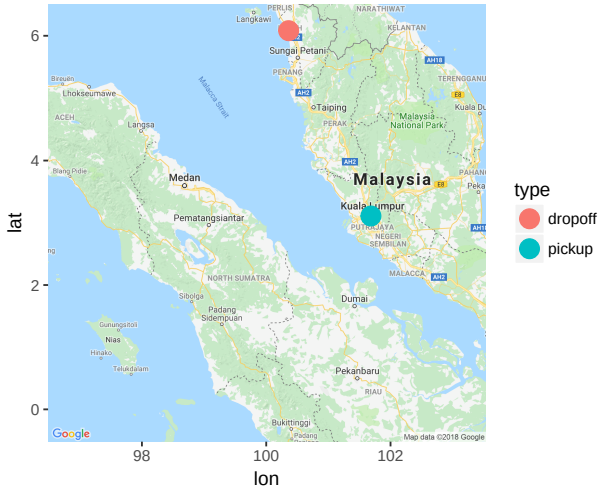
이동경로에 대한 분석 gif로 제시해도 좋을 것 같음.

Kuala Lumpur -> Kedah

```
example_id <- location_1 %>%
  dplyr::filter(booking_id == 11935)

map_1 <- get_map(location = c(lon = 100, lat = 3), zoom = 7, maptype = "roadmap")
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=3,100&zoom=7&size=640x640&scale=2&map
```

```
direction_example <- ggmap(map_1) + geom_point(aes(x = lng, y = lat, col = type), data = example_id, si

direction_example
```

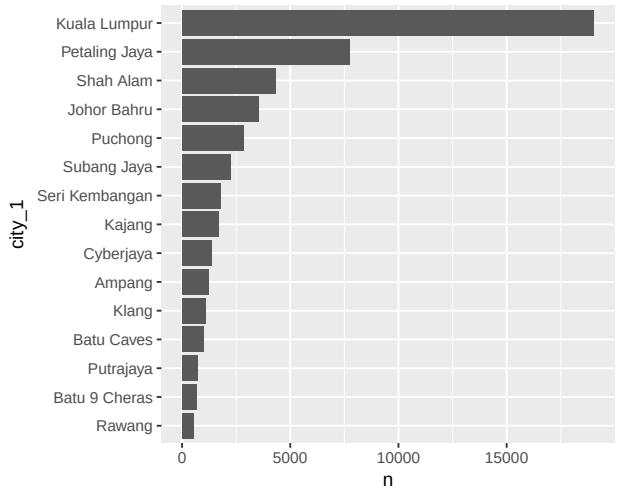


3. 도시 사용량 순위

순위권 도시를 뽑아서 그 도시 위주로 우선 분석이 필요할 것 같음

```
#City
## 456개 도시에서 사용

city <- location_1 %>%
  dplyr::count(city) %>%
  dplyr::arrange(desc(n)) %>%
  dplyr::mutate(
    city_1 = fct_reorder(city, n, "mean")
  )

## 사용량 Top 15 city

ggplot(city[1:15,], aes(x = city_1, y = n)) +
```

```
  geom_col() +
  coord_flip()
```

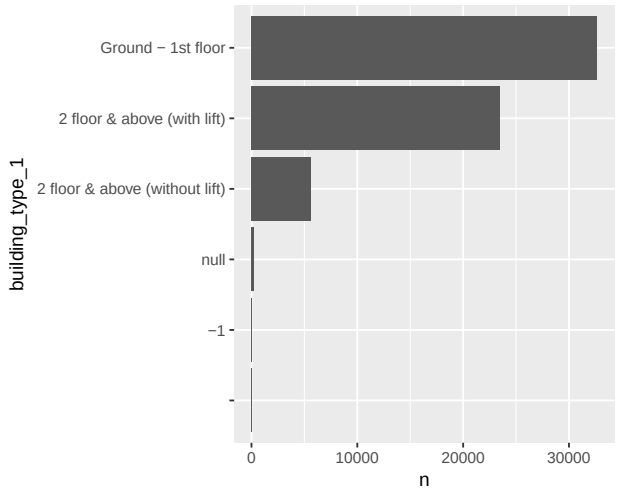


4. Buliding_type

크게 세종류

300 개 정도 오염된 데이터(결측값 및 알수없는 변수명)

```
building <- location_1 %>%
  dplyr::count(building_type) %>%
  dplyr::arrange(desc(n)) %>%
  dplyr::mutate(
    building_type_1 = fct_reorder(building_type, n, "mean")
  )

ggplot(building, aes(x = building_type_1, y = n)) +
  geom_col() +
  coord_flip()
```



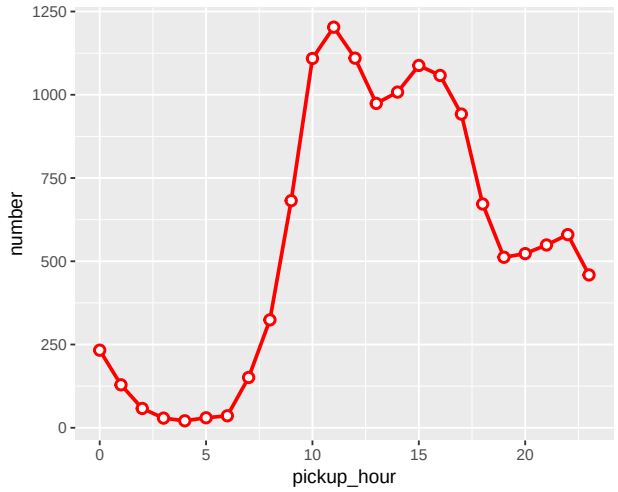
abandoned 수가 비중이 많다. -> 문제

점심시간 사용비중이 높ㄷ

```
table(booking$booking_status)
```

```
##
## abandoned completed   on-hold   pending  rejected
##     11199     13480        43        25      3760
```

```
booking_time <- booking %>%
  dplyr::filter(booking_status == "completed") %>%
  dplyr::group_by(hour(booking_datetime)) %>%
  dplyr::count()

colnames(booking_time) <- c("pickup_hour", "number")

ggplot(booking_time, aes(x = pickup_hour, y = number)) +
  geom_line(size = 1,color = 'red') +
  geom_point(shape = 21,
             size = 2,
             stroke = 1.2,
             color = 'red',
             fill = 'white')
```



11, 12월 사용량이 많음

```
booking_month <- booking %>%
  dplyr::filter(booking_status == "completed") %>%
  dplyr::group_by(month(booking_datetime)) %>%
  dplyr::count()

colnames(booking_month) <- c("pickup_month", "number")

ggplot(booking_month, aes(x = as.factor(pickup_month), y = number)) +
  geom_col()
```