

dissertation

2024-06-04

packages

data preparation

folder path

```
path_to_folder <- "/Users/wangyue/Desktop/dissertation_data/"
```

shenzhen shapefile

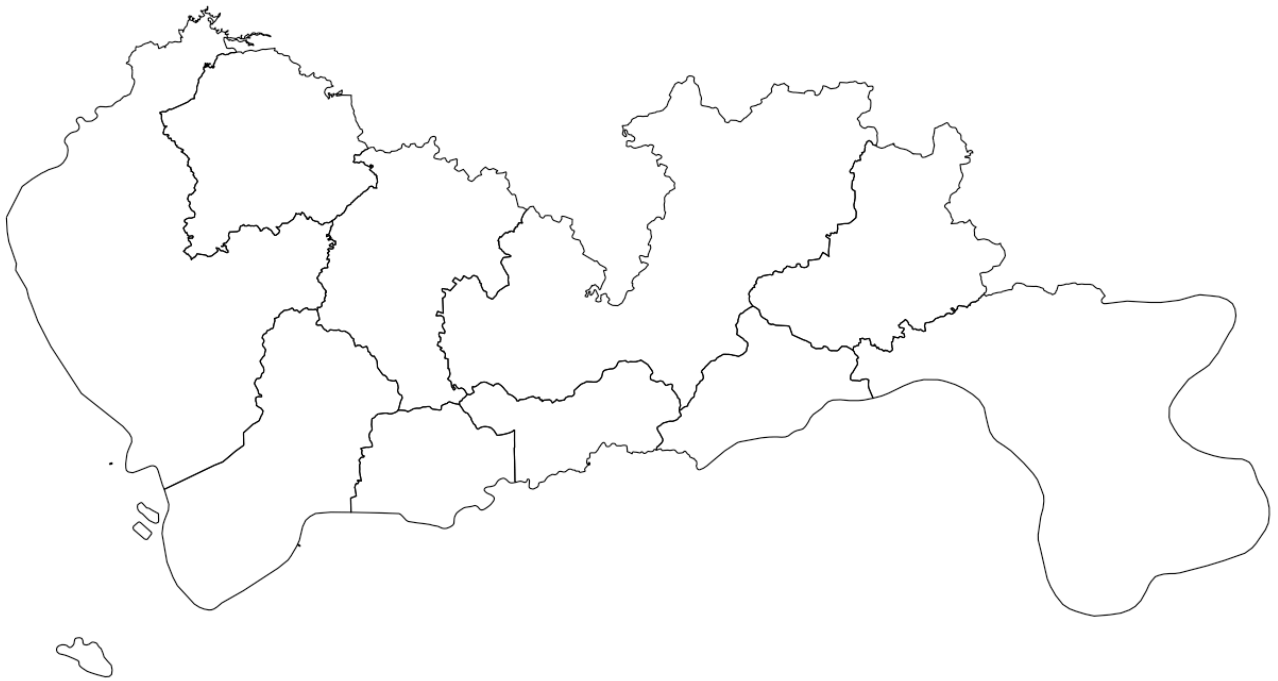
```
shenzhen <- st_read(paste0(path_to_folder, "/shenzhen/440300.shp") )
```

```
## Reading layer `440300' from data source
##   `/Users/wangyue/Desktop/dissertation_data/shenzhen/440300.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 9 features and 3 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: 113.7463 ymin: 22.39929 xmax: 114.624 ymax: 22.86472
## CRS:            NA
```

```
shenzhen <- shenzhen %>%
  dplyr::select(id, name, geometry)

shenzhen <- st_set_crs(shenzhen, 4547)

ggplot(data = shenzhen) +
  geom_sf(fill = NA, color = "black") +
  theme_minimal() +
  theme(line = element_blank(),
        axis.text=element_blank())
```



poi data (Cainiao stations)

import shapefile

filter poi data

```

poi2013 <- poi2013 %>%
  dplyr::select(name, adname, adcode, geometry)

poi2014 <- poi2014 %>%
  dplyr::select(NAME, COUNTY, CODE, geometry) %>%
  rename(name = NAME, adname = COUNTY, adcode = CODE)

poi2015 <- poi2015 %>%
  dplyr::select(NAME, COUNTY, CODE, geometry) %>%
  rename(name = NAME, adname = COUNTY, adcode = CODE)

poi2016 <- poi2016 %>%
  dplyr::select(NAME, COUNTY, CODE, geometry) %>%
  rename(name = NAME, adname = COUNTY, adcode = CODE)

poi2017 <- poi2017 %>%
  dplyr::select(name, adname, adcode, geometry) %>%
  filter(str_detect(name, "菜鸟") | str_detect(name, "自提柜") | str_detect(name,
"快递柜") | str_detect(name, "驿站"))

poi2018 <- poi2018 %>%
  filter(str_detect(name, "菜鸟") | str_detect(name, "自提柜") | str_detect(name, "快
递柜") | str_detect(name, "驿站") | str_detect(name, "智能柜")) %>%
  dplyr::select(name, adname, adcode, geometry)

poi2019 <- poi2019 %>%
  filter(str_detect(name, "菜鸟") | str_detect(name, "自提柜") | str_detect(name, "快
递柜") | str_detect(name, "驿站") | str_detect(name, "智能柜")) %>%
  dplyr::select(name, adname, adcode, geometry)

poi2020 <- poi2020 %>%
  filter(str_detect(name, "菜鸟") | str_detect(name, "自提柜") | str_detect(name, "快
递柜") | str_detect(name, "驿站") | str_detect(name, "智能柜") | str_detect(name, "速
递易柜")) %>%
  dplyr::select(name, adname, adcode, geometry)

poi2021 <- poi2021 %>%
  filter(str_detect(name, "菜鸟") | str_detect(name, "自提柜") | str_detect(name, "快
递柜") | str_detect(name, "驿站") | str_detect(name, "智能柜") | str_detect(name, "速
递易柜") | str_detect(name, "快递超市")) %>%
  dplyr::select(name, adname, adcode, geometry)

poi2022 <- poi2022 %>%
  filter(str_detect(name, "菜鸟") | str_detect(name, "自提柜") | str_detect(name, "快
递柜") | str_detect(name, "驿站") | str_detect(name, "智能柜") | str_detect(name, "速
递易柜") | str_detect(name, "快递超市") | str_detect(name, "快递点") | str_detect(nam
e, "送货柜")) %>%
  dplyr::select(name, adname, adcode, geometry)

```

nighttime light data

###Li(2020) ####2013

```

Li2013 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2013.tif"))
Li2013 <- crop(Li2013, shenzhen)
Li2013 <- mask(Li2013, shenzhen)

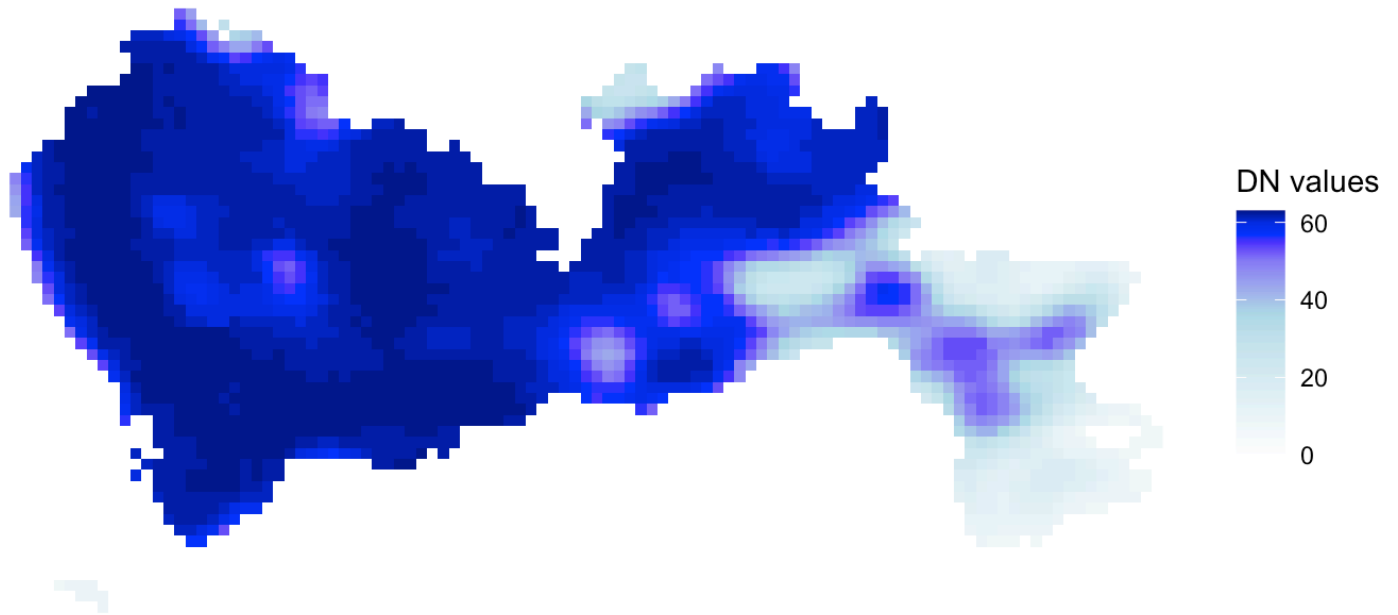
Li2013_df <- as.data.frame(Li2013, xy = TRUE, na.rm = TRUE)
names(Li2013_df)[3] <- "value"

Li2013_df <- Li2013_df %>%
  filter(value > 7)

Li2013_map <- ggplot() +
  geom_raster(data = Li2013_df, aes(x = x, y = y, fill = value)) +
  scale_fill_gradientn(
    colors = c("white", "lightblue", "blue", "darkblue"),
    values = scales::rescale(c(0, 30, 50, 60, 63)),
    limits = c(min(0), max(Li2013_df$value, na.rm = TRUE)),
    oob = scales::squish
  ) +
  coord_sf() +
  theme_void() +
  theme(
    legend.position = "right",
    plot.title = element_text(hjust = 0.5) # Center the title
  ) +
  labs(fill = "DN values") +
  ggtitle("Nighttime Light Value in 2013")
Li2013_map

```

Nighttime Light Value in 2013



####2014

```
Li2014 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2014.tif"))
Li2014 <- crop(Li2014, shenzhen)
Li2014 <- mask(Li2014, shenzhen)

Li2014_df <- as.data.frame(Li2014, xy = TRUE, na.rm = TRUE)
names(Li2014_df)[3] <- "value"
Li2014_df <- Li2014_df %>%
  filter(value > 7)
```

####2015

```
Li2015 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2015.tif"))
Li2015 <- crop(Li2015, shenzhen)
Li2015 <- mask(Li2015, shenzhen)

Li2015_df <- as.data.frame(Li2015, xy = TRUE, na.rm = TRUE)
names(Li2015_df)[3] <- "value"
Li2015_df <- Li2015_df %>%
  filter(value > 7)
```

2016

```

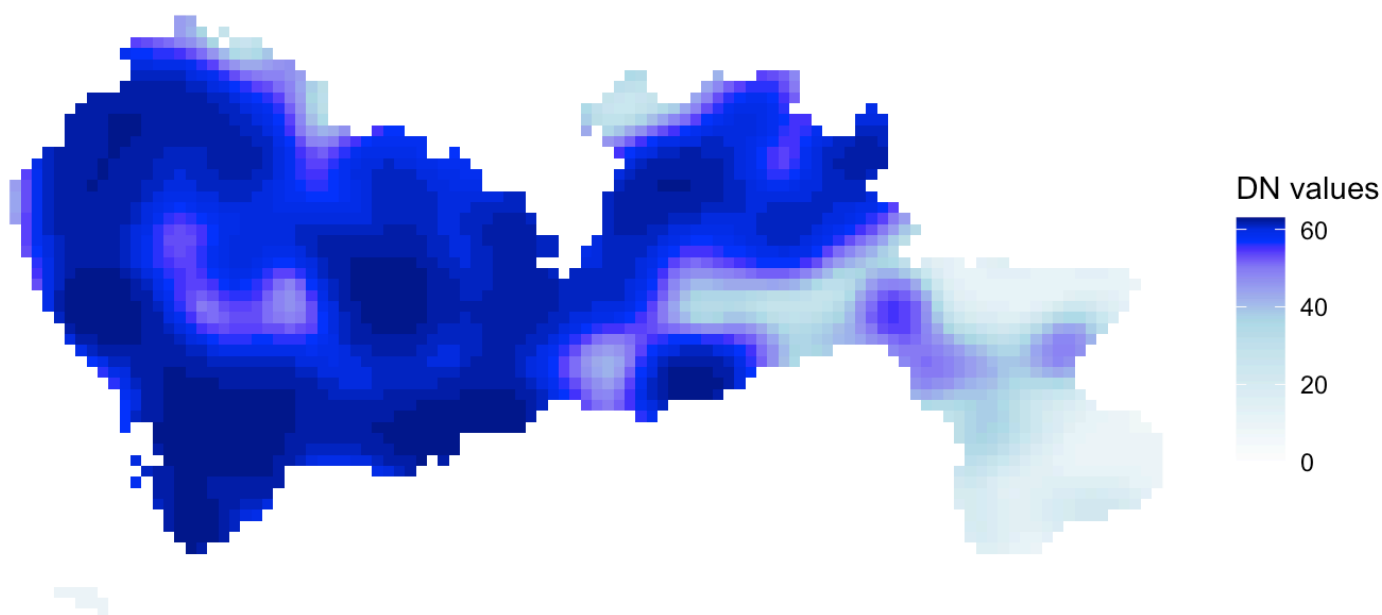
Li2016 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2016.tif"))
Li2016 <- crop(Li2016, shenzhen)
Li2016 <- mask(Li2016, shenzhen)

Li2016_df <- as.data.frame(Li2016, xy = TRUE, na.rm = TRUE)
names(Li2016_df)[3] <- "value"
Li2016_df <- Li2016_df %>%
  filter(value > 7)

Li2016_map <- ggplot() +
  geom_raster(data = Li2016_df, aes(x = x, y = y, fill = value)) +
  scale_fill_gradientn(
    colors = c("white", "lightblue", "blue", "darkblue"),
    values = scales::rescale(c(0, 30, 50, 60, 63)),
    limits = c(min(0), max(Li2016_df$value, na.rm = TRUE)),
    oob = scales::squish
  ) +
  coord_sf() +
  theme_void() +
  theme(
    legend.position = "right",
    plot.title = element_text(hjust = 0.5)
  ) +
  labs(fill = "DN values") +
  ggtitle("Nighttime Light Value in 2016")
Li2016_map

```

Nighttime Light Value in 2016



####2017

```
Li2017 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2017.tif"))
Li2017 <- crop(Li2017, shenzhen)
Li2017 <- mask(Li2017, shenzhen)

Li2017_df <- as.data.frame(Li2017, xy = TRUE, na.rm = TRUE)
names(Li2017_df)[3] <- "value"
Li2017_df <- Li2017_df %>%
  filter(value > 7)
```

####2018

```
Li2018 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2018.tif"))
Li2018 <- crop(Li2018, shenzhen)
Li2018 <- mask(Li2018, shenzhen)

Li2018_df <- as.data.frame(Li2018, xy = TRUE, na.rm = TRUE)
names(Li2018_df)[3] <- "value"
Li2018_df <- Li2018_df %>%
  filter(value > 7)
```

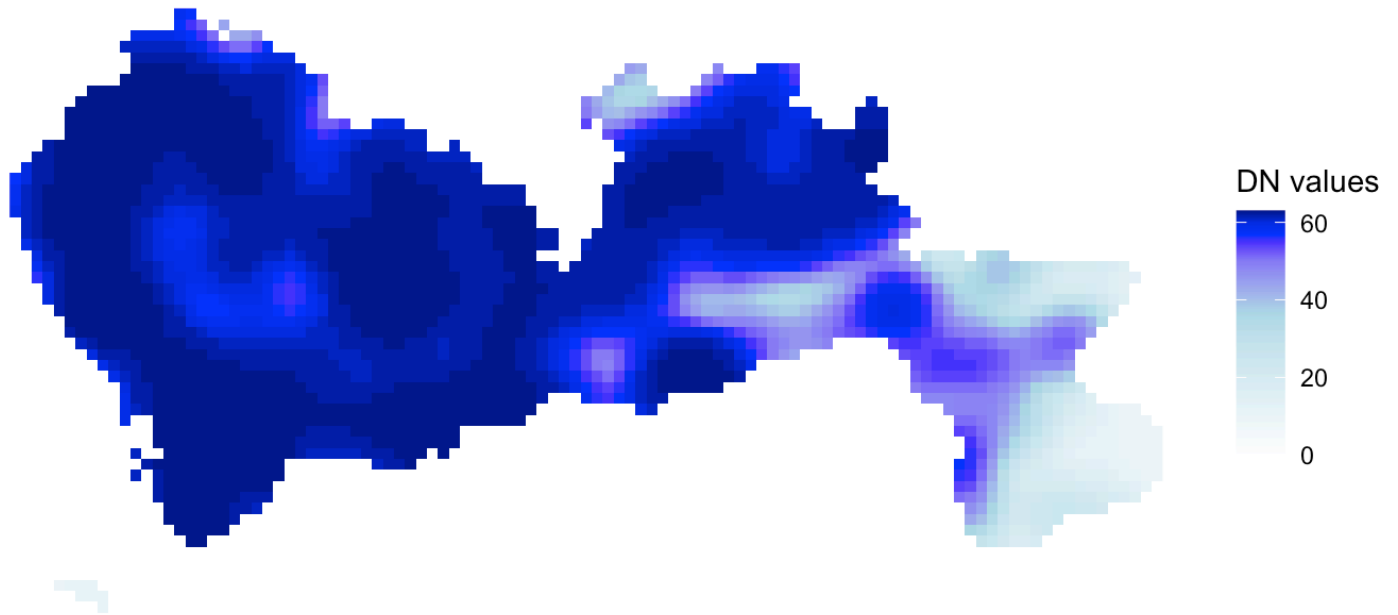
2019

```
Li2019 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2019.tif"))
Li2019 <- crop(Li2019, shenzhen)
Li2019 <- mask(Li2019, shenzhen)

Li2019_df <- as.data.frame(Li2019, xy = TRUE, na.rm = TRUE)
names(Li2019_df)[3] <- "value"
Li2019_df <- Li2019_df %>%
  filter(value > 7)

Li2019_map <- ggplot() +
  geom_raster(data = Li2019_df, aes(x = x, y = y, fill = value)) +
  scale_fill_gradientn(
    colors = c("white", "lightblue", "blue", "darkblue"),
    values = scales::rescale(c(0, 30, 50, 60, 63)),
    limits = c(min(0), max(Li2019_df$value, na.rm = TRUE)),
    oob = scales::squish
  ) +
  coord_sf() +
  theme_void() +
  theme(
    legend.position = "right",
    plot.title = element_text(hjust = 0.5) # Center the title
  ) +
  labs(fill = "DN values") +
  ggtitle("Nighttime Light Value in 2019")
Li2019_map
```

Nighttime Light Value in 2019



####2020

```
Li2020 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2020.tif"))
Li2020 <- crop(Li2020, shenzhen)
Li2020 <- mask(Li2020, shenzhen)

Li2020_df <- as.data.frame(Li2020, xy = TRUE, na.rm = TRUE)
names(Li2020_df)[3] <- "value"
Li2020_df <- Li2020_df %>%
  filter(value > 7)
```

####2021

```
Li2021 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2021.tif"))
Li2021 <- crop(Li2021, shenzhen)
Li2021 <- mask(Li2021, shenzhen)

Li2021_df <- as.data.frame(Li2021, xy = TRUE, na.rm = TRUE)
names(Li2021_df)[3] <- "value"
Li2021_df <- Li2021_df %>%
  filter(value > 7)
```

2022


```

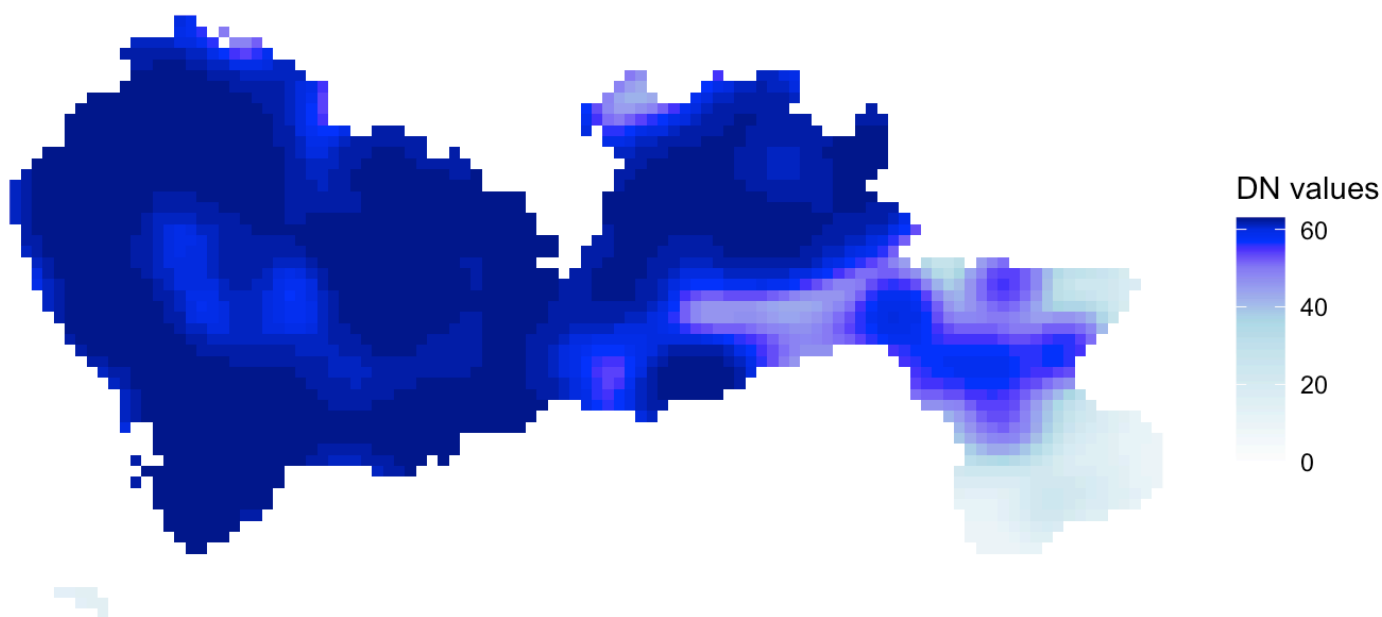
Li2022 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2022.tif"))
Li2022 <- crop(Li2022, shenzhen)
Li2022 <- mask(Li2022, shenzhen)

Li2022_df <- as.data.frame(Li2022, xy = TRUE, na.rm = TRUE)
names(Li2022_df)[3] <- "value"
Li2022_df <- Li2022_df %>%
  filter(value > 7)

Li2022_map <- ggplot() +
  geom_raster(data = Li2022_df, aes(x = x, y = y, fill = value)) +
  scale_fill_gradientn(
    colors = c("white", "lightblue", "blue", "darkblue"),
    values = scales::rescale(c(0, 30, 50, 60, 63)),
    limits = c(min(0), max(Li2022_df$value, na.rm = TRUE)),
    oob = scales::squish
  ) +
  coord_sf() +
  theme_void() +
  theme(
    legend.position = "right",
    plot.title = element_text(hjust = 0.5)
  ) +
  labs(fill = "DN values") +
  ggtitle("Nighttime Light Value in 2022")
Li2022_map

```

Nighttime Light Value in 2022



measurement of parcel stations

nearest distance to parcel stations

```
Li2013 <- rasterToPoints(Li2013, spatial = TRUE)
Li2013 <- st_as_sf(Li2013)
Li2013 <- st_transform(Li2013, crs = 4547)
names(Li2013)[1] <- "value"
Li2013 <- Li2013 %>%
  filter(value > 7)
```

```
Li2014 <- rasterToPoints(Li2014, spatial = TRUE)
Li2014 <- st_as_sf(Li2014)
Li2014 <- st_transform(Li2014, crs = 4547)
names(Li2014)[1] <- "value"
Li2014 <- Li2014 %>%
  filter(value > 7)
```

```
Li2015 <- rasterToPoints(Li2015, spatial = TRUE)
Li2015 <- st_as_sf(Li2015)
Li2015 <- st_transform(Li2015, crs = 4547)
names(Li2015)[1] <- "value"
Li2015 <- Li2015 %>%
  filter(value > 7)
```

```
Li2016 <- rasterToPoints(Li2016, spatial = TRUE)
Li2016 <- st_as_sf(Li2016)
Li2016 <- st_transform(Li2016, crs = 4547)
names(Li2016)[1] <- "value"
Li2016 <- Li2016 %>%
  filter(value > 7)
```

```
Li2017 <- rasterToPoints(Li2017, spatial = TRUE)
Li2017 <- st_as_sf(Li2017)
Li2017 <- st_transform(Li2017, crs = 4547)
names(Li2017)[1] <- "value"
Li2017 <- Li2017 %>%
  filter(value > 7)
```

```
Li2018 <- rasterToPoints(Li2018, spatial = TRUE)
Li2018 <- st_as_sf(Li2018)
Li2018 <- st_transform(Li2018, crs = 4547)
names(Li2018)[1] <- "value"
Li2018 <- Li2018 %>%
  filter(value > 7)
```

```
Li2019 <- rasterToPoints(Li2019, spatial = TRUE)
Li2019 <- st_as_sf(Li2019)
Li2019 <- st_transform(Li2019, crs = 4547)
names(Li2019)[1] <- "value"
Li2019 <- Li2019 %>%
```

```
filter(value > 7)
```

```
Li2020 <- rasterToPoints(Li2020, spatial = TRUE)
Li2020 <- st_as_sf(Li2020)
Li2020 <- st_transform(Li2020, crs = 4547)
names(Li2020)[1] <- "value"
Li2020 <- Li2020 %>%
  filter(value > 7)
```

```
Li2021 <- rasterToPoints(Li2021, spatial = TRUE)
Li2021 <- st_as_sf(Li2021)
Li2021 <- st_transform(Li2021, crs = 4547)
names(Li2021)[1] <- "value"
Li2021 <- Li2021 %>%
  filter(value > 7)
```

```
Li2022 <- rasterToPoints(Li2022, spatial = TRUE)
Li2022 <- st_as_sf(Li2022)
Li2022 <- st_transform(Li2022, crs = 4547)
names(Li2022)[1] <- "value"
Li2022 <- Li2022 %>%
  filter(value > 7)
```

```
distances <- st_distance(Li2013, poi2013, by_element = FALSE)
Li2013$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2014, poi2014, by_element = FALSE)
Li2014$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2015, poi2015, by_element = FALSE)
Li2015$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2016, poi2016, by_element = FALSE)
Li2016$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2017, poi2017, by_element = FALSE)
Li2017$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2018, poi2018, by_element = FALSE)
Li2018$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2019, poi2019, by_element = FALSE)
Li2019$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2020, poi2020, by_element = FALSE)
Li2020$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2021, poi2021, by_element = FALSE)
Li2021$nearest_station_distance <- apply(distances, 1, min)
```

```
distances <- st_distance(Li2022, poi2022, by_element = FALSE)
Li2022$nearest_station_distance <- apply(distances, 1, min)
```

```
Li2013 <- Li2013 %>%
  rename(ntl = value)
Li2013$year <- 2013

Li2014 <- Li2014 %>%
  rename(ntl = value)
Li2014$year <- 2014

Li2015 <- Li2015 %>%
  rename(ntl = value)
Li2015$year <- 2015

Li2016 <- Li2016 %>%
  rename(ntl = value)
Li2016$year <- 2016

Li2017 <- Li2017 %>%
  rename(ntl = value)
Li2017$year <- 2017

Li2018 <- Li2018 %>%
  rename(ntl = value)
Li2018$year <- 2018

Li2019 <- Li2019 %>%
  rename(ntl = value)
Li2019$year <- 2019

Li2020 <- Li2020 %>%
  rename(ntl = value)
Li2020$year <- 2020

Li2021 <- Li2021 %>%
  rename(ntl = value)
Li2021$year <- 2021

Li2022 <- Li2022 %>%
  rename(ntl = value)
Li2022$year <- 2022
```

```
nearest_dis <- bind_rows(Li2013, Li2014, Li2015, Li2016, Li2017, Li2018, Li2019, Li
2020, Li2021, Li2022)
```

buffer

re-load dataset

```
Li2013 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2013.tif"))
Li2013 <- crop(Li2013, shenzhen)
Li2013 <- mask(Li2013, shenzhen)
Li2013 <- rasterToPoints(Li2013, spatial = TRUE)
Li2013 <- st_as_sf(Li2013)
```

```

Li2013 <- st_transform(Li2013, crs = 4547)
names(Li2013)[1] <- "value"
Li2013 <- Li2013 %>%
  filter(value > 7)

Li2014 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2014.tif"))
Li2014 <- crop(Li2014, shenzhen)
Li2014 <- mask(Li2014, shenzhen)
Li2014 <- rasterToPoints(Li2014, spatial = TRUE)
Li2014 <- st_as_sf(Li2014)
Li2014 <- st_transform(Li2014, crs = 4547)
names(Li2014)[1] <- "value"
Li2014 <- Li2014 %>%
  filter(value > 7)

Li2015 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2015.tif"))
Li2015 <- crop(Li2015, shenzhen)
Li2015 <- mask(Li2015, shenzhen)
Li2015 <- rasterToPoints(Li2015, spatial = TRUE)
Li2015 <- st_as_sf(Li2015)
Li2015 <- st_transform(Li2015, crs = 4547)
names(Li2015)[1] <- "value"
Li2015 <- Li2015 %>%
  filter(value > 7)

Li2016 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2016.tif"))
Li2016 <- crop(Li2016, shenzhen)
Li2016 <- mask(Li2016, shenzhen)
Li2016 <- rasterToPoints(Li2016, spatial = TRUE)
Li2016 <- st_as_sf(Li2016)
Li2016 <- st_transform(Li2016, crs = 4547)
names(Li2016)[1] <- "value"
Li2016 <- Li2016 %>%
  filter(value > 7)

Li2017 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2017.tif"))
Li2017 <- crop(Li2017, shenzhen)
Li2017 <- mask(Li2017, shenzhen)
Li2017 <- rasterToPoints(Li2017, spatial = TRUE)
Li2017 <- st_as_sf(Li2017)
Li2017 <- st_transform(Li2017, crs = 4547)
names(Li2017)[1] <- "value"
Li2017 <- Li2017 %>%
  filter(value > 7)

Li2018 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2018.tif"))
Li2018 <- crop(Li2018, shenzhen)
Li2018 <- mask(Li2018, shenzhen)
Li2018 <- rasterToPoints(Li2018, spatial = TRUE)
Li2018 <- st_as_sf(Li2018)
Li2018 <- st_transform(Li2018, crs = 4547)
names(Li2018)[1] <- "value"
Li2018 <- Li2018 %>%
  filter(value > 7)

```

```
Li2019 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2019.tif"))
Li2019 <- crop(Li2019, shenzhen)
Li2019 <- mask(Li2019, shenzhen)
Li2019 <- rasterToPoints(Li2019, spatial = TRUE)
Li2019 <- st_as_sf(Li2019)
Li2019 <- st_transform(Li2019, crs = 4547)
names(Li2019)[1] <- "value"
Li2019 <- Li2019 %>%
  filter(value > 7)

Li2020 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2020.tif"))
Li2020 <- crop(Li2020, shenzhen)
Li2020 <- mask(Li2020, shenzhen)
Li2020 <- rasterToPoints(Li2020, spatial = TRUE)
Li2020 <- st_as_sf(Li2020)
Li2020 <- st_transform(Li2020, crs = 4547)
names(Li2020)[1] <- "value"
Li2020 <- Li2020 %>%
  filter(value > 7)

Li2021 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2021.tif"))
Li2021 <- crop(Li2021, shenzhen)
Li2021 <- mask(Li2021, shenzhen)
Li2021 <- rasterToPoints(Li2021, spatial = TRUE)
Li2021 <- st_as_sf(Li2021)
Li2021 <- st_transform(Li2021, crs = 4547)
names(Li2021)[1] <- "value"
Li2021 <- Li2021 %>%
  filter(value > 7)

Li2022 <- raster(paste0(path_to_folder, "/nighttime_light/Li2020/2022.tif"))
Li2022 <- crop(Li2022, shenzhen)
Li2022 <- mask(Li2022, shenzhen)
Li2022 <- rasterToPoints(Li2022, spatial = TRUE)
Li2022 <- st_as_sf(Li2022)
Li2022 <- st_transform(Li2022, crs = 4547)
names(Li2022)[1] <- "value"
Li2022 <- Li2022 %>%
  filter(value > 7)
```

```
Li2013 <- Li2013 %>%
  rename(ntl = value)
Li2013$year <- 2013

Li2014 <- Li2014 %>%
  rename(ntl = value)
Li2014$year <- 2014

Li2015 <- Li2015 %>%
  rename(ntl = value)
Li2015$year <- 2015

Li2016 <- Li2016 %>%
  rename(ntl = value)
Li2016$year <- 2016

Li2017 <- Li2017 %>%
  rename(ntl = value)
Li2017$year <- 2017

Li2018 <- Li2018 %>%
  rename(ntl = value)
Li2018$year <- 2018

Li2019 <- Li2019 %>%
  rename(ntl = value)
Li2019$year <- 2019

Li2020 <- Li2020 %>%
  rename(ntl = value)
Li2020$year <- 2020

Li2021 <- Li2021 %>%
  rename(ntl = value)
Li2021$year <- 2021

Li2022 <- Li2022 %>%
  rename(ntl = value)
Li2022$year <- 2022
```

k = 1

```

buffer_dist <- 1000
Li2013_k1 <- st_buffer(Li2013, dist = buffer_dist)
Li2013_k1$buffer_id <- 1:nrow(Li2013_k1)
buffer_cainiao <- st_intersects(Li2013_k1, poi2013) %>%lengths()
Li2013_k1$parcel_station_count <- buffer_cainiao

Li2014_k1 <- st_buffer(Li2014, dist = buffer_dist)
Li2014_k1$buffer_id <- 1:nrow(Li2014_k1)
buffer_cainiao <- st_intersects(Li2014_k1, poi2014) %>% lengths()
Li2014_k1$parcel_station_count <- buffer_cainiao

Li2015_k1 <- st_buffer(Li2015, dist = buffer_dist)
Li2015_k1$buffer_id <- 1:nrow(Li2015_k1)
buffer_cainiao <- st_intersects(Li2015_k1, poi2015) %>%lengths()
Li2015_k1$parcel_station_count <- buffer_cainiao

Li2016_k1 <- st_buffer(Li2016, dist = buffer_dist)
Li2016_k1$buffer_id <- 1:nrow(Li2016_k1)
buffer_cainiao <- st_intersects(Li2016_k1, poi2016) %>%lengths()
Li2016_k1$parcel_station_count <- buffer_cainiao

Li2017_k1 <-st_buffer(Li2017, dist = buffer_dist)
Li2017_k1$buffer_id <- 1:nrow(Li2017_k1)
buffer_cainiao <- st_intersects(Li2017_k1, poi2017) %>%lengths()
Li2017_k1$parcel_station_count <- buffer_cainiao

Li2018_k1 <- st_buffer(Li2018, dist = buffer_dist)
Li2018_k1$buffer_id <- 1:nrow(Li2018_k1)
buffer_cainiao <- st_intersects(Li2018_k1, poi2018) %>%lengths()
Li2018_k1$parcel_station_count <- buffer_cainiao

Li2019_k1 <- st_buffer(Li2019, dist = buffer_dist)
Li2019_k1$buffer_id <- 1:nrow(Li2019_k1)
buffer_cainiao <- st_intersects(Li2019_k1, poi2019) %>%lengths()
Li2019_k1$parcel_station_count <- buffer_cainiao

Li2020_k1 <- st_buffer(Li2020, dist = buffer_dist)
Li2020_k1$buffer_id <- 1:nrow(Li2020_k1)
buffer_cainiao <- st_intersects(Li2020_k1, poi2020) %>%lengths()
Li2020_k1$parcel_station_count <- buffer_cainiao

Li2021_k1 <- st_buffer(Li2021, dist = buffer_dist)
Li2021_k1$buffer_id <- 1:nrow(Li2021_k1)
buffer_cainiao <- st_intersects(Li2021_k1, poi2021) %>%lengths()
Li2021_k1$parcel_station_count <- buffer_cainiao

Li2022_k1 <- st_buffer(Li2022, dist = buffer_dist)
Li2022_k1$buffer_id <- 1:nrow(Li2022_k1)
buffer_cainiao <- st_intersects(Li2022_k1, poi2022) %>%lengths()
Li2022_k1$parcel_station_count <- buffer_cainiao

```

k = 3


```

buffer_dist <- 3000
Li2013_k3 <- st_buffer(Li2013, dist = buffer_dist)
Li2013_k3$buffer_id <- 1:nrow(Li2013_k3)
buffer_cainiao <- st_intersects(Li2013_k3, poi2013) %>%lengths()
Li2013_k3$parcel_station_count <- buffer_cainiao

Li2014_k3 <- st_buffer(Li2014, dist = buffer_dist)
Li2014_k3$buffer_id <- 1:nrow(Li2014_k3)
buffer_cainiao <- st_intersects(Li2014_k3, poi2014) %>%lengths()
Li2014_k3$parcel_station_count <- buffer_cainiao

Li2015_k3 <- st_buffer(Li2015, dist = buffer_dist)
Li2015_k3$buffer_id <- 1:nrow(Li2015_k3)
buffer_cainiao <- st_intersects(Li2015_k3, poi2015) %>%lengths()
Li2015_k3$parcel_station_count <- buffer_cainiao

Li2016_k3 <- st_buffer(Li2016, dist = buffer_dist)
Li2016_k3$buffer_id <- 1:nrow(Li2016_k3)
buffer_cainiao <- st_intersects(Li2016_k3, poi2016) %>%lengths()
Li2016_k3$parcel_station_count <- buffer_cainiao

Li2017_k3 <- st_buffer(Li2017, dist = buffer_dist)
Li2017_k3$buffer_id <- 1:nrow(Li2017_k3)
buffer_cainiao <- st_intersects(Li2017_k3, poi2017) %>%lengths()
Li2017_k3$parcel_station_count <- buffer_cainiao

Li2018_k3 <- st_buffer(Li2018, dist = buffer_dist)
Li2018_k3$buffer_id <- 1:nrow(Li2018_k3)
buffer_cainiao <- st_intersects(Li2018_k3, poi2018) %>%lengths()
Li2018_k3$parcel_station_count <- buffer_cainiao

Li2019_k3 <- st_buffer(Li2019, dist = buffer_dist)
Li2019_k3$buffer_id <- 1:nrow(Li2019_k3)
buffer_cainiao <- st_intersects(Li2019_k3, poi2019) %>%lengths()
Li2019_k3$parcel_station_count <- buffer_cainiao

Li2020_k3 <- st_buffer(Li2020, dist = buffer_dist)
Li2020_k3$buffer_id <- 1:nrow(Li2020_k3)
buffer_cainiao <- st_intersects(Li2020_k3, poi2020) %>%lengths()
Li2020_k3$parcel_station_count <- buffer_cainiao

Li2021_k3 <- st_buffer(Li2021, dist = buffer_dist)
Li2021_k3$buffer_id <- 1:nrow(Li2021_k3)
buffer_cainiao <- st_intersects(Li2021_k3, poi2021) %>%lengths()
Li2021_k3$parcel_station_count <- buffer_cainiao

Li2022_k3 <- st_buffer(Li2022, dist = buffer_dist)
Li2022_k3$buffer_id <- 1:nrow(Li2022_k3)
buffer_cainiao <- st_intersects(Li2022_k3, poi2022) %>%lengths()
Li2022_k3$parcel_station_count <- buffer_cainiao

```

k = 5

```

buffer_dist <- 5000
Li2013_k5 <- st_buffer(Li2013, dist = buffer_dist)
Li2013_k5$buffer_id <- 1:nrow(Li2013_k5)
buffer_cainiao <- st_intersects(Li2013_k5, poi2013) %>%lengths()
Li2013_k5$parcel_station_count <- buffer_cainiao

Li2014_k5 <- st_buffer(Li2014, dist = buffer_dist)
Li2014_k5$buffer_id <- 1:nrow(Li2014_k5)
buffer_cainiao <- st_intersects(Li2014_k5, poi2014) %>%lengths()
Li2014_k5$parcel_station_count <- buffer_cainiao

Li2015_k5 <- st_buffer(Li2015, dist = buffer_dist)
Li2015_k5$buffer_id <- 1:nrow(Li2015_k5)
buffer_cainiao <- st_intersects(Li2015_k5, poi2015) %>%lengths()
Li2015_k5$parcel_station_count <- buffer_cainiao

Li2016_k5 <- st_buffer(Li2016, dist = buffer_dist)
Li2016_k5$buffer_id <- 1:nrow(Li2016_k5)
buffer_cainiao <- st_intersects(Li2016_k5, poi2016) %>%lengths()
Li2016_k5$parcel_station_count <- buffer_cainiao

Li2017_k5 <- st_buffer(Li2017, dist = buffer_dist)
Li2017_k5$buffer_id <- 1:nrow(Li2017_k5)
buffer_cainiao <- st_intersects(Li2017_k5, poi2017) %>%lengths()
Li2017_k5$parcel_station_count <- buffer_cainiao

Li2018_k5 <- st_buffer(Li2018, dist = buffer_dist)
Li2018_k5$buffer_id <- 1:nrow(Li2018_k5)
buffer_cainiao <- st_intersects(Li2018_k5, poi2018) %>%lengths()
Li2018_k5$parcel_station_count <- buffer_cainiao

Li2019_k5 <- st_buffer(Li2019, dist = buffer_dist)
Li2019_k5$buffer_id <- 1:nrow(Li2019_k5)
buffer_cainiao <- st_intersects(Li2019_k5, poi2019) %>%lengths()
Li2019_k5$parcel_station_count <- buffer_cainiao

Li2020_k5 <- st_buffer(Li2020, dist = buffer_dist)
Li2020_k5$buffer_id <- 1:nrow(Li2020_k5)
buffer_cainiao <- st_intersects(Li2020_k5, poi2020) %>%lengths()
Li2020_k5$parcel_station_count <- buffer_cainiao

Li2021_k5 <- st_buffer(Li2021, dist = buffer_dist)
Li2021_k5$buffer_id <- 1:nrow(Li2021_k5)
buffer_cainiao <- st_intersects(Li2021_k5, poi2021) %>%lengths()
Li2021_k5$parcel_station_count <- buffer_cainiao

Li2022_k5 <- st_buffer(Li2022, dist = buffer_dist)
Li2022_k5$buffer_id <- 1:nrow(Li2022_k5)
buffer_cainiao <- st_intersects(Li2022_k5, poi2022) %>%lengths()
Li2022_k5$parcel_station_count <- buffer_cainiao

```

combine dataset

```

buffer_k1 <- bind_rows(Li2013_k1, Li2014_k1, Li2015_k1, Li2016_k1, Li2017_k1, Li2018_k1, Li2019_k1, Li2020_k1, Li2021_k1, Li2022_k1)
buffer_k1$nearest_station_distance <- nearest_dis$nearest_station_distance

buffer_k3 <- bind_rows(Li2013_k3, Li2014_k3, Li2015_k3, Li2016_k3, Li2017_k3, Li2018_k3, Li2019_k3, Li2020_k3, Li2021_k3, Li2022_k3)

buffer_k5 <- bind_rows(Li2013_k5, Li2014_k5, Li2015_k5, Li2016_k5, Li2017_k5, Li2018_k5, Li2019_k5, Li2020_k5, Li2021_k5, Li2022_k5)

```

OLS

```

OLS_k1 <- lm(ntl ~ parcel_station_count, data = buffer_k1)
OLS_k3 <- lm(ntl ~ parcel_station_count, data = buffer_k3)
OLS_k5 <- lm(ntl ~ parcel_station_count, data = buffer_k5)

OLS_nearest <- lm(ntl ~ nearest_station_distance, data = nearest_dis)

summary(OLS_nearest)

```

```

##
## Call:
## lm(formula = ntl ~ nearest_station_distance, data = nearest_dis)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -50.428  -0.514    2.331    3.914   45.046
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.111e+01  6.953e-02   878.9  <2e-16 ***
## nearest_station_distance -2.296e-03  1.273e-05  -180.3  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.573 on 26946 degrees of freedom
## Multiple R-squared:  0.5469, Adjusted R-squared:  0.5469
## F-statistic: 3.252e+04 on 1 and 26946 DF, p-value: < 2.2e-16

```

```
summary(OLS_k1)
```

```
##
## Call:
## lm(formula = ntl ~ parcel_station_count, data = buffer_k1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -44.602  -1.602   5.990   8.398  10.398
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    52.60228    0.09175   573.3  <2e-16 ***
## parcel_station_count  0.48150    0.01082    44.5  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.73 on 26946 degrees of freedom
## Multiple R-squared:  0.06846, Adjusted R-squared:  0.06842
## F-statistic: 1980 on 1 and 26946 DF, p-value: < 2.2e-16
```

```
summary(OLS_k3)
```

```
##
## Call:
## lm(formula = ntl ~ parcel_station_count, data = buffer_k3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -45.357  -2.300   5.019   8.643  11.700
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    51.300377    0.095256  538.55  <2e-16 ***
## parcel_station_count  0.097943    0.001628   60.17  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.35 on 26946 degrees of freedom
## Multiple R-squared:  0.1184, Adjusted R-squared:  0.1184
## F-statistic: 3620 on 1 and 26946 DF, p-value: < 2.2e-16
```

```
summary(OLS_k5)
```

```
##
## Call:
## lm(formula = ntl ~ parcel_station_count, data = buffer_k5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -45.306  -2.562   4.380   9.043  12.484
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.052e+01  9.793e-02  515.82  <2e-16 ***
## parcel_station_count 4.650e-02  6.938e-04   67.03  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.17 on 26946 degrees of freedom
## Multiple R-squared:  0.1429, Adjusted R-squared:  0.1429
## F-statistic:  4493 on 1 and 26946 DF,  p-value: < 2.2e-16
```

panel data

```
panel_data_k1 <- pdata.frame(buffer_k1, index = c("buffer_id", "year"))
panel_data_k3 <- pdata.frame(buffer_k3, index = c("buffer_id", "year"))
panel_data_k5 <- pdata.frame(buffer_k5, index = c("buffer_id", "year"))
```

fixed effects

```
panel_k1 <- plm(ntl ~ parcel_station_count + factor(year), data = panel_data_k1, model = "within")
panel_k3 <- plm(ntl ~ parcel_station_count + factor(year), data = panel_data_k3, model = "within")
panel_k5 <- plm(ntl ~ parcel_station_count + factor(year), data = panel_data_k5, model = "within")

panel_nearest <- plm(ntl ~ nearest_station_distance + factor(year), data = panel_data_k1, model = "within")

summary(panel_k1) summary(panel_k3) summary(panel_k5) summary(panel_nearest)
```

limitation: parcel station count is too small for the panel data analysis. Not suitable in this model.

```
panel_nearest <- plm(ntl ~ nearest_station_distance + factor(year), data = panel_data_k1, model = "within")
summary(panel_nearest)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = ntl ~ nearest_station_distance + factor(year),
##      data = panel_data_k1, model = "within")
##
## Unbalanced Panel: n = 2708, T = 7-10, N = 26948
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -45.14972  -1.59720   0.16489   1.85319   37.78862
##
## Coefficients:
##
##              Estimate Std. Error t-value Pr(>|t|)
## nearest_station_distance -1.0863e-03  1.3652e-05 -79.5701 < 2.2e-16 ***
## factor(year)2014          9.4486e-01  1.7058e-01   5.5392 3.070e-08 ***
## factor(year)2015          1.7042e+00  1.7162e-01   9.9300 < 2.2e-16 ***
## factor(year)2016         -2.1336e+00  1.6857e-01 -12.6566 < 2.2e-16 ***
## factor(year)2017         -2.5099e+00  1.7054e-01 -14.7180 < 2.2e-16 ***
## factor(year)2018         -2.3023e+00  1.7158e-01 -13.4182 < 2.2e-16 ***
## factor(year)2019          6.9160e-01  1.6974e-01   4.0744 4.629e-05 ***
## factor(year)2020         -1.1338e+00  1.7248e-01  -6.5735 5.016e-11 ***
## factor(year)2021          1.3050e+00  1.7049e-01   7.6545 2.013e-14 ***
## factor(year)2022          1.4047e+00  1.7054e-01   8.2367 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:      1254700
## Residual Sum of Squares: 927120
## R-Squared:      0.2611
## Adj. R-Squared: 0.17825
## F-statistic: 856.201 on 10 and 24230 DF, p-value: < 2.22e-16
```

count data

Negative Binomial Regression

```
mean_value <- mean(panel_data_k1$parcel_station_count, na.rm = TRUE)
variance_value <- var(panel_data_k1$parcel_station_count, na.rm = TRUE)

mean_value
```

```
## [1] 3.491057
```

```
variance_value
```

```
## [1] 59.71391
```

```
summary(panel_data_k1$parcel_station_count)
```

```
## total sum of squares: 1609111
##           id           time
## 0.3716619 0.1468209
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.000   0.000   0.000   3.491   3.000   93.000
```

```
# Negative Binomial Regression
```

```
negbin_k1 <- fenegbin(ntl ~ parcel_station_count | year, data = panel_data_k1)
negbin_k3 <- fenegbin(ntl ~ parcel_station_count | year, data = panel_data_k3)
negbin_k5 <- fenegbin(ntl ~ parcel_station_count | year, data = panel_data_k5)
```

```
summary(negbin_k1)
```

```
## ML estimation, family = Negative Binomial, Dep. Var.: ntl
## Observations: 26,948
## Fixed-effects: year: 10
## Standard-errors: Clustered (year)
##              Estimate Std. Error z value  Pr(>|z|)
## parcel_station_count 0.008498    0.001801 4.71836 2.3775e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Over-dispersion parameter: theta = 11.96117
## Log-Likelihood: -115,261.6   Adj. Pseudo R2: 0.005149
##              BIC: 230,635.5    Squared Cor.: 0.063532
```

```
summary(negbin_k3)
```

```
## ML estimation, family = Negative Binomial, Dep. Var.: ntl
## Observations: 26,948
## Fixed-effects: year: 10
## Standard-errors: Clustered (year)
##              Estimate Std. Error z value  Pr(>|z|)
## parcel_station_count 0.002074    0.000454 4.5707 4.8609e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Over-dispersion parameter: theta = 12.63011
## Log-Likelihood: -114,741.5   Adj. Pseudo R2: 0.009637
##              BIC: 229,595.3    Squared Cor.: 0.107563
```

```
summary(negbin_k5)
```

```
## ML estimation, family = Negative Binomial, Dep. Var.: ntl
## Observations: 26,948
## Fixed-effects: year: 10
## Standard-errors: Clustered (year)
##
##              Estimate Std. Error z value   Pr(>|z|)
## parcel_station_count 0.001127    0.000238 4.74311 2.1047e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Over-dispersion parameter: theta = 13.17101
## Log-Likelihood: -114,348.3   Adj. Pseudo R2: 0.013031
##              BIC:   228,808.9       Squared Cor.: 0.140856
```

pooled OLS

```
buffer_k1$buffer_id <- as.factor(buffer_k1$buffer_id)
buffer_k1$year <- as.factor(buffer_k1$year)

buffer_k3$buffer_id <- as.factor(buffer_k3$buffer_id)
buffer_k3$year <- as.factor(buffer_k3$year)

buffer_k5$buffer_id <- as.factor(buffer_k5$buffer_id)
buffer_k5$year <- as.factor(buffer_k5$year)

# Pooled OLS
pooled_ols_lm_nearest <- lm(ntl ~ nearest_station_distance, data = buffer_k1)
pooled_ols_lm_k1 <- lm(ntl ~ parcel_station_count, data = buffer_k1)
pooled_ols_lm_k3 <- lm(ntl ~ parcel_station_count, data = buffer_k3)
pooled_ols_lm_k5 <- lm(ntl ~ parcel_station_count, data = buffer_k5)

pooled_ols_plm_k1 <- plm(ntl ~ parcel_station_count, data = panel_data_k1, model =
"pooling")
pooled_ols_plm_k3 <- plm(ntl ~ parcel_station_count, data = panel_data_k3, model =
"pooling")
pooled_ols_plm_k5 <- plm(ntl ~ parcel_station_count, data = panel_data_k5, model =
"pooling")

summary(pooled_ols_lm_nearest)
```



```
##
## Call:
## lm(formula = ntl ~ nearest_station_distance, data = buffer_k1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -50.428  -0.514   2.331   3.914  45.046
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.111e+01  6.953e-02   878.9  <2e-16 ***
## nearest_station_distance -2.296e-03  1.273e-05  -180.3  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.573 on 26946 degrees of freedom
## Multiple R-squared:  0.5469, Adjusted R-squared:  0.5469
## F-statistic: 3.252e+04 on 1 and 26946 DF, p-value: < 2.2e-16
```

```
summary(pooled_ols_lm_k1)
```

```
##
## Call:
## lm(formula = ntl ~ parcel_station_count, data = buffer_k1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -44.602  -1.602   5.990   8.398  10.398
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      52.60228    0.09175   573.3  <2e-16 ***
## parcel_station_count  0.48150    0.01082   44.5  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.73 on 26946 degrees of freedom
## Multiple R-squared:  0.06846, Adjusted R-squared:  0.06842
## F-statistic: 1980 on 1 and 26946 DF, p-value: < 2.2e-16
```

```
summary(pooled_ols_plm_k1)
```

```
## Pooling Model
##
## Call:
## plm(formula = ntl ~ parcel_station_count, data = panel_data_k1,
##      model = "pooling")
##
## Unbalanced Panel: n = 2708, T = 7-10, N = 26948
##
## Residuals:
##      Min.   1st Qu.   Median   3rd Qu.    Max.
## -44.6023  -1.6023   5.9902   8.3977  10.3977
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)    52.602283   0.091748  573.33 < 2.2e-16 ***
## parcel_station_count  0.481496   0.010820   44.50 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    5449400
## Residual Sum of Squares: 5076300
## R-Squared:    0.068458
## Adj. R-Squared: 0.068423
## F-statistic: 1980.23 on 1 and 26946 DF, p-value: < 2.22e-16
```

```
summary(pooled_ols_plm_k3)
```

```
## Pooling Model
##
## Call:
## plm(formula = ntl ~ parcel_station_count, data = panel_data_k3,
##      model = "pooling")
##
## Unbalanced Panel: n = 2708, T = 7-10, N = 26948
##
## Residuals:
##      Min.   1st Qu.   Median   3rd Qu.    Max.
## -45.3572  -2.3004   5.0189   8.6428  11.6996
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)    51.3003773   0.0952561  538.552 < 2.2e-16 ***
## parcel_station_count  0.0979432   0.0016279   60.166 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    5449400
## Residual Sum of Squares: 4804000
## R-Squared:    0.11843
## Adj. R-Squared: 0.1184
## F-statistic: 3619.9 on 1 and 26946 DF, p-value: < 2.22e-16
```

```
summary(pooled_ols_plm_k5)
```

```
## Pooling Model
##
## Call:
## plm(formula = ntl ~ parcel_station_count, data = panel_data_k5,
##      model = "pooling")
##
## Unbalanced Panel: n = 2708, T = 7-10, N = 26948
##
## Residuals:
##      Min.   1st Qu.   Median   3rd Qu.    Max.
## -45.3061  -2.5624   4.3800   9.0425  12.4841
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)    5.0516e+01  9.7932e-02  515.823 < 2.2e-16 ***
## parcel_station_count 4.6504e-02  6.9377e-04   67.032 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    5449400
## Residual Sum of Squares: 4670600
## R-Squared:    0.14292
## Adj. R-Squared: 0.14289
## F-statistic: 4493.26 on 1 and 26946 DF, p-value: < 2.22e-16
```

Lag effects

```

panel_lag_data_k1 <- as.data.table(panel_data_k1)
panel_lag_data_k3 <- as.data.table(panel_data_k3)
panel_lag_data_k5 <- as.data.table(panel_data_k5)

panel_lag_data_nearest <- panel_lag_data_k1[order(buffer_id, year)]
panel_lag_data_nearest[, nearest_lag1 := shift(nearest_station_distance, 1, type =
"lag"), by = buffer_id]
panel_lag_data_nearest[, nearest_lag2 := shift(nearest_station_distance, 2, type =
"lag"), by = buffer_id]
panel_lag_data_nearest[, nearest_lag3 := shift(nearest_station_distance, 3, type =
"lag"), by = buffer_id]

panel_lag_data_k1 <- panel_lag_data_k1[order(buffer_id, year)]
panel_lag_data_k1[, Station_lag1 := shift(parcel_station_count, 1, type = "lag"), b
y = buffer_id]
panel_lag_data_k1[, Station_lag2 := shift(parcel_station_count, 2, type = "lag"), b
y = buffer_id]
panel_lag_data_k1[, Station_lag3 := shift(parcel_station_count, 3, type = "lag"), b
y = buffer_id]

panel_lag_data_k3 <- panel_lag_data_k3[order(buffer_id, year)]
panel_lag_data_k3[, Station_lag1 := shift(parcel_station_count, 1, type = "lag"), b
y = buffer_id]
panel_lag_data_k3[, Station_lag2 := shift(parcel_station_count, 2, type = "lag"), b
y = buffer_id]
panel_lag_data_k3[, Station_lag3 := shift(parcel_station_count, 3, type = "lag"), b
y = buffer_id]

panel_lag_data_k5 <- panel_lag_data_k5[order(buffer_id, year)]
panel_lag_data_k5[, Station_lag1 := shift(parcel_station_count, 1, type = "lag"), b
y = buffer_id]
panel_lag_data_k5[, Station_lag2 := shift(parcel_station_count, 2, type = "lag"), b
y = buffer_id]
panel_lag_data_k5[, Station_lag3 := shift(parcel_station_count, 3, type = "lag"), b
y = buffer_id]

```

```

# remove geometry
panel_lag_data_nearest <- panel_lag_data_nearest %>%
  dplyr::select(-geometry)
panel_lag_data_k1 <- panel_lag_data_k1 %>%
  dplyr::select(-geometry)
panel_lag_data_k3 <- panel_lag_data_k3 %>%
  dplyr::select(-geometry)
panel_lag_data_k5 <- panel_lag_data_k5 %>%
  dplyr::select(-geometry)

# pdata.frame
panel_lag_data_nearest <- pdata.frame(panel_lag_data_nearest, index = c("buffer_i
d", "year"))
panel_lag_data_k1 <- pdata.frame(panel_lag_data_k1, index = c("buffer_id", "year"))
panel_lag_data_k3 <- pdata.frame(panel_lag_data_k3, index = c("buffer_id", "year"))
panel_lag_data_k5 <- pdata.frame(panel_lag_data_k5, index = c("buffer_id", "year"))

```

```
# negative binomial with fixed effects
lag_nearest <- plm(ntl ~ nearest_station_distance + nearest_lag1 + nearest_lag2 + nearest_lag3 + factor(year), data = panel_lag_data_nearest)

lag_k1 <- fenegbin(ntl ~ parcel_station_count + Station_lag1 + Station_lag2 + Station_lag3 | year, data = panel_lag_data_k1)
```

```
## NOTE: 8,124 observations removed because of NA values (RHS: 8,124).
```

```
lag_k3 <- fenegbin(ntl ~ parcel_station_count + Station_lag1 + Station_lag2 + Station_lag3 | year, data = panel_lag_data_k3)
```

```
## NOTE: 8,124 observations removed because of NA values (RHS: 8,124).
```

```
lag_k5 <- fenegbin(ntl ~ parcel_station_count + Station_lag1 + Station_lag2 + Station_lag3 | year, data = panel_lag_data_k5)
```

```
## NOTE: 8,124 observations removed because of NA values (RHS: 8,124).
```

```
summary(lag_nearest)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = ntl ~ nearest_station_distance + nearest_lag1 +
##       nearest_lag2 + nearest_lag3 + factor(year), data = panel_lag_data_nearest)
##
## Unbalanced Panel: n = 2708, T = 4-7, N = 18824
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -47.156351  -1.414782   0.055407   1.842355   31.197122
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## nearest_station_distance -3.0392e-03  3.2267e-05 -94.1896 < 2.2e-16 ***
## nearest_lag1             1.2030e-05  1.7082e-05   0.7042 0.4812992
## nearest_lag2            -3.3000e-04  1.3660e-05 -24.1572 < 2.2e-16 ***
## nearest_lag3            -4.8637e-04  1.3115e-05 -37.0862 < 2.2e-16 ***
## factor(year)2017         -1.8662e+00  1.4849e-01 -12.5673 < 2.2e-16 ***
## factor(year)2018         -3.2791e+00  1.5746e-01 -20.8247 < 2.2e-16 ***
## factor(year)2019         -5.6994e-01  1.5653e-01  -3.6410 0.0002724 ***
## factor(year)2020         -4.1024e+00  1.6000e-01 -25.6395 < 2.2e-16 ***
## factor(year)2021         -1.4916e+00  1.6171e-01  -9.2234 < 2.2e-16 ***
## factor(year)2022         -1.2103e+00  1.5957e-01  -7.5847 3.513e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:      705120
## Residual Sum of Squares: 389480
## R-Squared:      0.44763
## Adj. R-Squared: 0.35445
## F-statistic: 1305.21 on 10 and 16106 DF, p-value: < 2.22e-16
```

```
summary(lag_k1)
```

```
## ML estimation, family = Negative Binomial, Dep. Var.: ntl
## Observations: 18,824
## Fixed-effects: year: 7
## Standard-errors: Clustered (year)
##              Estimate Std. Error   z value   Pr(>|z|)
## parcel_station_count  0.009027   0.001765  5.114758 3.1414e-07 ***
## Station_lag1          0.001212   0.001504  0.806227 4.2011e-01
## Station_lag2          0.000680   0.001593  0.426733 6.6957e-01
## Station_lag3         -0.005304   0.003496 -1.517071 1.2925e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Over-dispersion parameter: theta = 14.35446
## Log-Likelihood: -79,529.4   Adj. Pseudo R2: 0.007658
##              BIC: 159,167.1   Squared Cor.: 0.090169
```

```
summary(lag_k3)
```

```
## ML estimation, family = Negative Binomial, Dep. Var.: ntl
## Observations: 18,824
## Fixed-effects: year: 7
## Standard-errors: Clustered (year)
##
##               Estimate Std. Error   z value   Pr(>|z|)
## parcel_station_count  0.002347   0.000417  5.626336 1.8408e-08 ***
## Station_lag1          0.000315   0.000337  0.933826 3.5039e-01
## Station_lag2          0.000199   0.000405  0.491487 6.2308e-01
## Station_lag3         -0.001738   0.000984 -1.766134 7.7373e-02 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Over-dispersion parameter: theta = 15.85489
## Log-Likelihood: -78,907.3   Adj. Pseudo R2: 0.01542
##               BIC: 157,922.9   Squared Cor.: 0.162409
```

```
summary(lag_k5)
```

```
## ML estimation, family = Negative Binomial, Dep. Var.: ntl
## Observations: 18,824
## Fixed-effects: year: 7
## Standard-errors: Clustered (year)
##
##               Estimate Std. Error   z value   Pr(>|z|)
## parcel_station_count  0.001252   0.000209  5.976907 2.2741e-09 ***
## Station_lag1          0.000178   0.000154  1.155415 2.4792e-01
## Station_lag2          0.000165   0.000206  0.801483 4.2285e-01
## Station_lag3         -0.000934   0.000497 -1.876605 6.0572e-02 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Over-dispersion parameter: theta = 17.14251
## Log-Likelihood: -78,435.9   Adj. Pseudo R2: 0.0213
##               BIC: 156,980.2   Squared Cor.: 0.213478
```