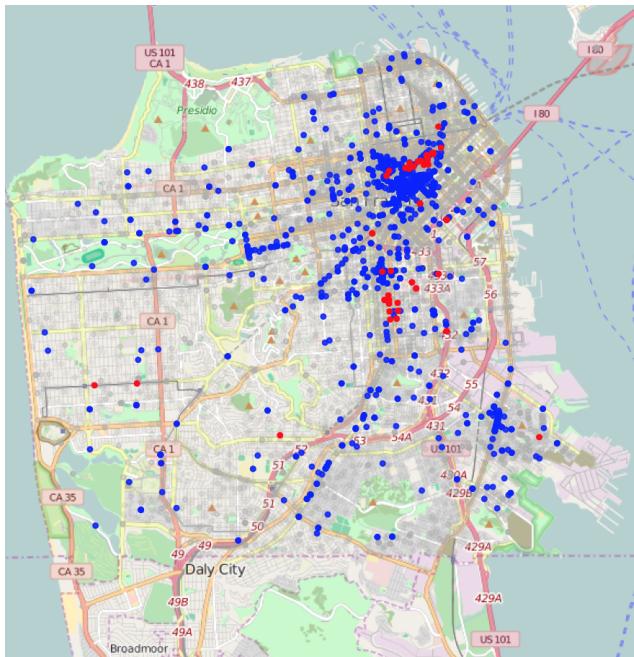


In both cities prostitution and drug-related crimes happens next to each other but not at the same places

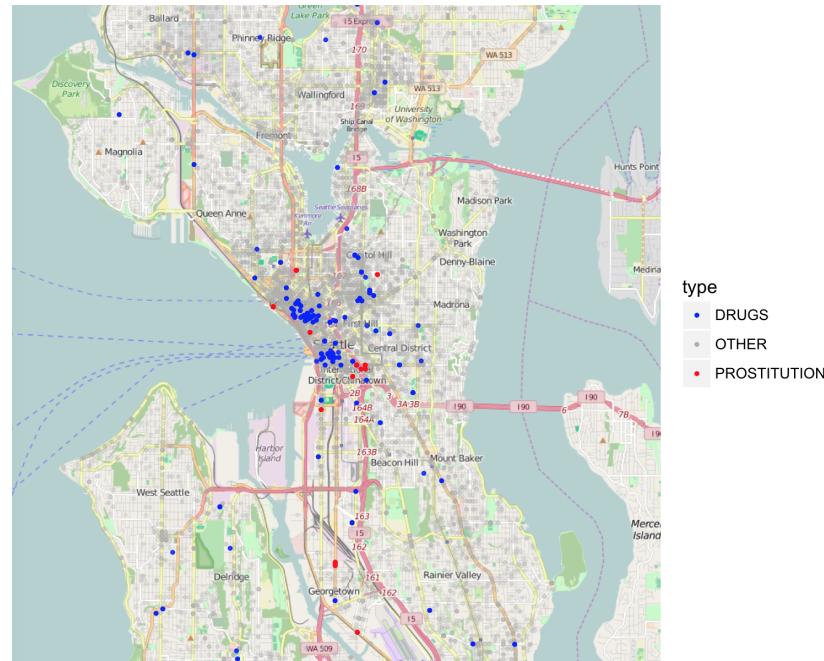
In attempt to verify conventional wisdom that drugs and prostitution always comes together I've decided to put them on the map of both San Francisco and Seattle and compare against all other types of incidents. On the maps below you may see that both prostitution and drug related crimes have higher concentration at the cities center. Even though clusters of both types are very close to each-other, it is clearly not the very same places.

It would be fair to say that concentration of both inspected types of crimes higher at areas which has higher incidents density across all types of crimes. But clusters does not overlap, that means that 'conventional wisdom' is not quite right.

San Francisco



Seattle



Approach

During the research major difficulty I've met was the fact that datasets available for San Francisco and Seattle has different structure of crimes classification. San Francisco's dataset gives 34 types of incidents, while Seattle offers 147. Seattle has 4 prostitution-related crime types and 28 drug related, while San Francisco has one type for each of them.

Taking R as a primary tool for analysis and 'ggplot' library as a visualization tool I've unified both datasets replacing crime types with 'PROSTITUTION' for all prostitution-related crimes, 'DRUGS' for every crime related to drugs or narcotics and 'OTHER' for all other types of crimes, as they were not in the focus of the study.

After that major drawback of plots I gain was that significantly large number of 'OTHER' incidents made concealed points I was interested in.

I've got much better picture after increasing transparency (reducing alpha channel to 0.1) of 'OTHER' points and sorting results by alpha channel (so points I was interested more appeared 'on-top' of background consisting of 'OTHER's).

Source code:

```
san <- read.csv("sanfrancisco_incidents_summer_2014.csv")
sea <- read.csv("seattle_incidents_summer_2014.csv")

library(dplyr)

# Normalize coords and Crime types
san_norm <- mutate(san,
  nx = X,
  ny = Y,
  type = ifelse(
    grepl('DRUG/NARCOTIC',Category), 'DRUGS',
    ifelse(grepl('PROSTITUTION',Category), 'PROSTITUTION', 'OTHER')),
  a = ifelse(
    grepl('DRUG/NARCOTIC',Category), 0.3,
    ifelse(grepl('PROSTITUTION',Category), 0.3, 0.005)),
  time_of_day = as.POSIXct(Time, format="%H:%M"),
  before_sunrise = ifelse(
    time_of_day>as.POSIXct('05:00', format='%H:%M'), TRUE, FALSE),
  after_dawn = ifelse(
    time_of_day>as.POSIXct('19:00', format='%H:%M'), TRUE, FALSE),
  night = (after_dawn | before_sunrise)
) %>%
  arrange(a)

sea_norm <- filter(sea, Latitude != 0, Longitude != 0 ) %>%
  mutate(
    ny = Latitude,
    nx = Longitude,
    type = ifelse(grepl('^NARC',Offense.Type), 'DRUGS',
      ifelse(grepl('^PROSTITUTION',Offense.Type), 'PROSTITUTION',
```

```

'OTHER')),
a = ifelse(grepl('^NARC',Offense.Type),0.3,
ifelse(grepl('^PROSTITUTION',Offense.Type),0.3,0.005)),
dt = parse_date_time(
Occurred.Date.or.Date.Range.Start, orders="mdy hms")
+ {ifelse(
grepl('PM$', Occurred.Date.or.Date.Range.Start), 12*3600, 0)
},
time_of_day =as.POSIXct(
strftime(dt, format='%I:%M'), format=%H:%M'),
before_sunrise = ifelse(
time_of_day>as.POSIXct('05:00', format=%H:%M'), TRUE, FALSE),
after_dawn = ifelse(
time_of_day>as.POSIXct('19:00', format=%H:%M'), TRUE, FALSE),
night = (after_dawn | before_sunrise)
) %>%
arrange(a)

require(ggmap)
sea_map <- qmap(location='seattle', zoom=12, source="osm", maptype = 'toner-lite')
sea_map + geom_point(data=sea_norm, aes(x=nx, y=ny, alpha=a, color=type), size=0.9) + guides(alpha=FALSE)
san_map <- qmap(location='san francisco', zoom=12, source="osm", maptype = 'toner-lite')
san_map + geom_point(data=san_norm, aes(x=nx, y=ny, alpha=a, color=type), size=0.9) + guides(alpha=FALSE)

```

Credit:

I'd like to thank you [Usman Khaliq](#) who's study taught me 'ggmap' library and helped make my visualizations became more aesthetically.
As well as the authors of 'ggmap' library: [D. Kahle and H. Wickham](#).