# Report1

## (1) Introduction

- map of tunisia
- star wars
- hannibal
- colesium
- lack of usage of alternative data in the developing world

## (2) Description of data

- Linkage data was collected from INS website
- INS website is very slow
- Planet didn't provide large enough coverage eg. at the country level
- Many sources for satellite imagery

## (3) Analysis of data quality

Some of the variable names are unclear, probably due to shoddy translation (i.e. what is the diff between 'Number of households having drinking water from the public source or source of water association' and 'Number of households having drinking water from the other public or private source'? - translation - stata
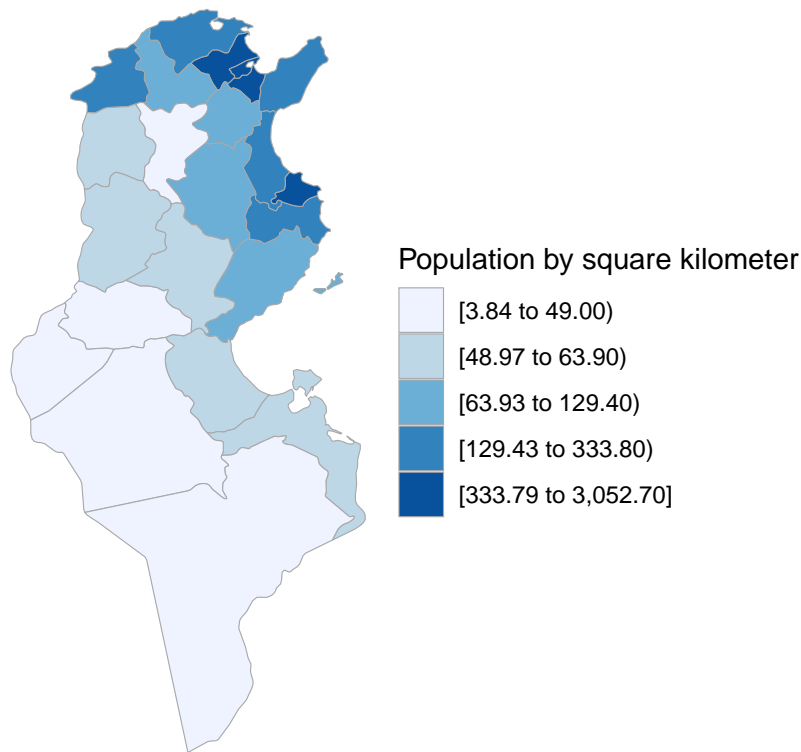
## (4) Main analysis (Exploratory Data Analysis)

Extensive data processing

### (4.a) A First Look at Tunisia

TODO: - supress the warning messages (add appropriate notation inside {r ... }) - remove Ariana from original data

```
#density <- read_excel("../data/intermediate/density.xlsx")
density <- readxl::read_xlsx("../data/intermediate/density.xlsx")
library(choroplethrAdmin1)
library(choroplethr)
admin1_choropleth("tunisia", density, num_colors = 5, title = "Population Density", legend = "Population
```

## Population Density



Population by square kilometer

- [3.84 to 49.00)
- [48.97 to 63.90)
- [63.93 to 129.40)
- [129.43 to 333.80)
- [333.79 to 3,052.70]

Comments: - coastal

### (4.b) Consumption by Governorates

TODO: - add after map bar plot or cleveland dot plot to show ranking of governorates by consumption

```r
consumption2015 <- readxl::read_xlsx("../data/intermediate/Enquête Consommation 2010 12_08_2018 10_49_0

names(consumption2015) <- consumption2015[1,] #copy 1st row
consumption2015 <- consumption2015[-1,] #remove 1st row from df

names(consumption2015)[1]<-"categories"

totalConsump15 <- filter(consumption2015, consumption2015$categories == "Total")
tidyConsump15 <- gather(totalConsump15, key="Governorates", value="Consumption")
tidyConsump15 <- tidyConsump15[-1,]
#tidyConsump15 <- filter(tidyConsump15, tidyConsump15$Governorates == "Total")

#removing Greater regions
tidyConsump15 <- filter(tidyConsump15, !Governorates %in% c("Great Tunis", "Governorate of Ariana", "Nor

#Note: Choroplethr does not include the governorate, "Governorate of Ariana"

#Rename governorate to match the ChoroplethrAdmin1 naming convention
tidyConsump15$Governorates <- c("gouvernorat de tunis", "gouvernorat de ben arous", "gouvernorat de la n

#as.numeric(as.character(exitDF$immigrant))
```
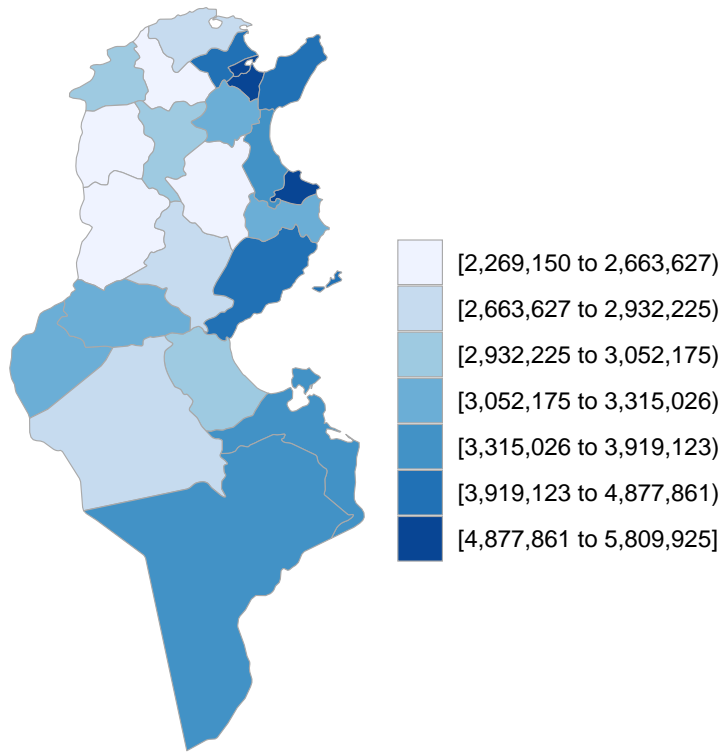
```r
df = data.frame(region=tidyConsump15$Governorates, value=as.numeric(as.character(tidyConsump15$Consumpt
```

```r
admin1_region_choropleth(df)
```



Comments: - show top 5 - focus on sfax


**(4.c) "[Namson] scatter plot lowess"**

```r
#dataset <- read_excel("../data/intermediate/dataset.xlsx")
dataset <- readxl::read_xlsx("../data/intermediate/dataset.xlsx")

ggplot(dataset, aes(population,mean)) +
    geom_point(color = "blue", size = 3) + geom_smooth(method = "lm", se = TRUE) + geom_label_repel(aes
```

Comments: - sfax outlier - general comment: eg. as population goes up luminsity goes up


**(4.d) A closer at outliers: Sfax**

```r
lum_del <- read_csv("../data/intermediate/tun_lum_delegation_93_13.csv")

## Parsed with column specification:
## cols(
##    .default = col_integer(),
##    GID_0 = col_character(),
##    NAME_0 = col_character(),
##    GID_1 = col_character(),
##    NAME_1 = col_character(),
```

```
##   NL_NAME_1 = col_character(),
##   GID_2 = col_character(),
##   NAME_2 = col_character(),
##   VARNAME_2 = col_character(),
##   NL_NAME_2 = col_character(),
##   TYPE_2 = col_character(),
##   ENGTYPE_2 = col_character(),
##   CC_2 = col_character(),
##   HASC_2 = col_character(),
##   `1992_mean` = col_double(),
##   `1992_media` = col_double(),
##   `1992_stdev` = col_double(),
##   `1992_varia` = col_double(),
##   `1993_mean` = col_double(),
##   `1993_media` = col_double(),
##   `1993_stdev` = col_double()
##   # ... with 81 more columns
## )

## See spec(...) for full column specifications.
```

```r
del_sfax <- filter(lum_del,lum_del$NAME_1 == "Sfax")
tbl_df(del_sfax)
```

```
## # A tibble: 15 x 277
##    GID_0 NAME_0 GID_1 NAME_1 NL_NAME_1 GID_2 NAME_2 VARNAME_2 NL_NAME_2
##    <chr> <chr>  <chr> <chr>  <chr>     <chr> <chr>  <chr>     <chr>
##  1 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Agareb <NA>      <NA>
##  2 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Bir A~ Bir Ali ~ <NA>
##  3 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ El Am~ Amra      <NA>
##  4 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ El Gh~ Gheriba   <NA>
##  5 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Hencha <NA>      <NA>
##  6 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Jeben~ <NA>      <NA>
##  7 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Kerke~ <NA>      <NA>
##  8 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Mahres El Mahres <NA>
##  9 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Menze~ <NA>      <NA>
## 10 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Sakie~ <NA>      <NA>
## 11 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Sakie~ <NA>      <NA>
## 12 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Sfax ~ Sfax Med~ <NA>
## 13 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Sfax ~ <NA>      <NA>
## 14 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Sfax ~ <NA>      <NA>
## 15 TUN   Tunis~ TUN.~ Sfax   <NA>      TUN.~ Skhira <NA>      <NA>
## # ... with 268 more variables: TYPE_2 <chr>, ENGTYPE_2 <chr>, CC_2 <chr>,
## #   HASC_2 <chr>, `1992_count` <int>, `1992_sum` <int>, `1992_mean` <dbl>,
## #   `1992_media` <dbl>, `1992_stdev` <dbl>, `1992_min` <int>,
## #   `1992_max` <int>, `1992_range` <int>, `1992_minor` <int>,
## #   `1992_major` <int>, `1992_varie` <int>, `1992_varia` <dbl>,
## #   `1993_count` <int>, `1993_sum` <int>, `1993_mean` <dbl>,
## #   `1993_media` <dbl>, `1993_stdev` <dbl>, `1993_min` <int>,
## #   `1993_max` <int>, `1993_range` <int>, `1993_minor` <int>,
## #   `1993_major` <int>, `1993_varie` <int>, `1993_varia` <dbl>,
## #   `1994_count` <int>, `1994_sum` <int>, `1994_mean` <dbl>,
## #   `1994_media` <dbl>, `1994_stdev` <dbl>, `1994_min` <int>,
## #   `1994_max` <int>, `1994_range` <int>, `1994_minor` <int>,
## #   `1994_major` <int>, `1994_varie` <int>, `1994_varia` <dbl>,
```

```
## #   `1995_count` <int>, `1995_sum` <int>, `1995_mean` <dbl>,
## #   `1995_media` <dbl>, `1995_stdev` <dbl>, `1995_min` <int>,
## #   `1995_max` <int>, `1995_range` <int>, `1995_minor` <int>,
## #   `1995_major` <int>, `1995_varie` <int>, `1995_varia` <dbl>,
## #   `1996_count` <int>, `1996_sum` <int>, `1996_mean` <dbl>,
## #   `1996_media` <dbl>, `1996_stdev` <dbl>, `1996_min` <int>,
## #   `1996_max` <int>, `1996_range` <int>, `1996_minor` <int>,
## #   `1996_major` <int>, `1996_varie` <int>, `1996_varia` <dbl>,
## #   `1997_count` <int>, `1997_sum` <int>, `1997_mean` <dbl>,
## #   `1997_media` <dbl>, `1997_stdev` <dbl>, `1997_min` <int>,
## #   `1997_max` <int>, `1997_range` <int>, `1997_minor` <int>,
## #   `1997_major` <int>, `1997_varie` <int>, `1997_varia` <dbl>,
## #   `1998_count` <int>, `1998_sum` <int>, `1998_mean` <dbl>,
## #   `1998_media` <dbl>, `1998_stdev` <dbl>, `1998_min` <int>,
## #   `1998_max` <int>, `1998_range` <int>, `1998_minor` <int>,
## #   `1998_major` <int>, `1998_varie` <int>, `1998_varia` <dbl>,
## #   `1999_count` <int>, `1999_sum` <int>, `1999_mean` <dbl>,
## #   `1999_media` <dbl>, `1999_stdev` <dbl>, `1999_min` <int>,
## #   `1999_max` <int>, `1999_range` <int>, `1999_minor` <int>,
## #   `1999_major` <int>, `1999_varie` <int>, `1999_varia` <dbl>, ...
```

```r
boxplot(del_sfax$`1992_mean`, horizontal = TRUE)
stripchart(del_sfax$`1992_mean`, col = "blue", pch = 21, add = TRUE, method = "jitter")
```
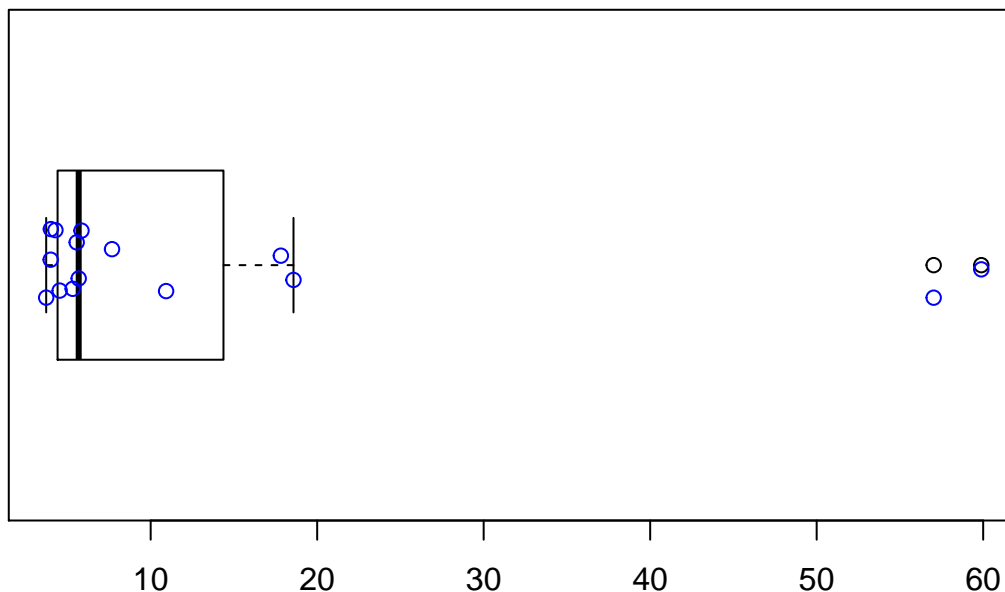


```r
ggplot(del_sfax, aes(del_sfax$NAME_2, del_sfax$`1992_mean`)) +
  geom_boxplot() +
  geom_jitter(alpha = 0.2, width = 0.1, color = "#926d25") +
  coord_flip() +labs(y = "Luminosity (mean)", x = "Sfax delegation") +
  ggtitle("What is wrong with Sfax?",
          subtitle = "A closer look at luminosity by delegations within Sfax") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35"))
```

**What is wrong with Sfax?**
A closer look at luminosity by delegations within Sfax
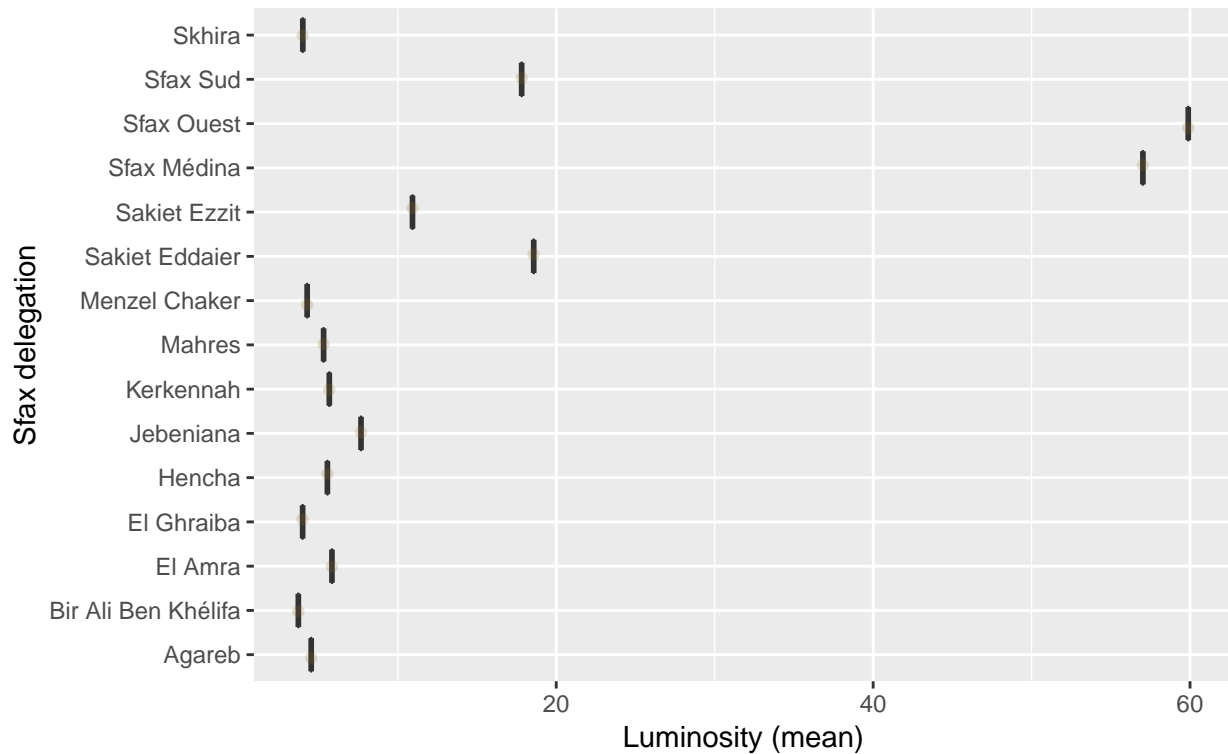
**(4.e) Exploring potential covariates of luminosity**

```r
appliances <- read_tun_data(here("data/raw",
                                 "Households by possession of electrical household equipements 11_18_20
  mutate(Region = str_replace(Region, "Governorate of ", "")) %>% # Standardize governorate names
  filter(Region != "Tunisia")


# household / population data
house_pop <- read_csv(here("data/intermediate",
                           "master_file.csv")) %>%
  rename(Region = governorate)

appliances <- inner_join(select(house_pop, Region, households), appliances)

# tidy
appliances <- appliances %>%
  gather(contains("_"), key="key", value="value") %>%
  separate(key, into=c("Appliance", "Measure"), sep="_") %>%
  spread(Appliance, value)

# we only want total
appliances_total <- appliances %>%
  filter(Measure == "Total") %>%
  select(-Measure)
```

```r
#fix the names
ap_names = c("Region", "households", "AC", "heat", "dishwasher", "stove", "fridge", "washing_machine")
names(appliances_total) <- ap_names

# cleaning up
appliances_total <- appliances_total %>%
  mutate_at(.vars = vars(-Region), funs(as.numeric)) %>%
  mutate(Region = str_replace(Region, "Governorate of ", ""))

appliances_tidy <- appliances_total %>%
  gather(-Region, -households, key="app_type", value="num_apps") %>%
  mutate(app_household = num_apps / households)

g <- ggplot(appliances_tidy, aes(x = app_household,
                                 y = fct_reorder2(Region, app_type, -app_household),
                                 color = app_type)) +
    geom_point() + ylab("") +
    ggtitle("Number of Appliances Per Household")
g
```

## (5) Executive summary (Presentation-style)

## (6) Interactive component

## (7) Conclusion

- Interested in looking at 2011 (year of the jasmine revolution) sicne the GDP went drastically down (Maybe less economic activity? Look into lumnisity?)
- Time series of luminosity per governorate
- !! Make a comment about coastal governorates
- https://academic.oup.com/eurpub/article/24/suppl_1/6/560448 - info on Choucha refugee crisis and 20% increase in Medeneine population
- Challenge: choroplethr naming for governorates is very specific and does not support french accent". Two regions are missing.
- Bin side - looked into curtomizing it but very time consuming