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

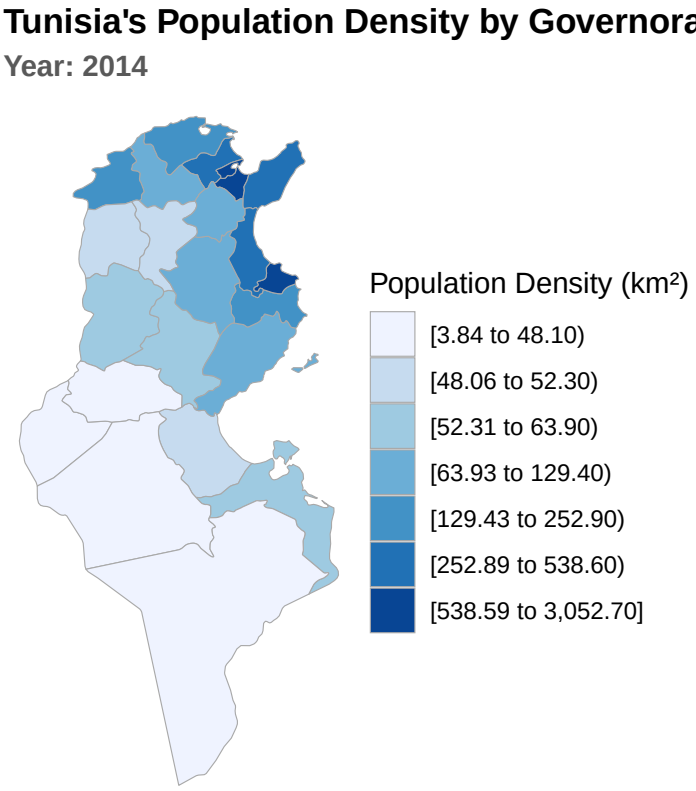
(4) Main analysis (Exploratory Data Analysis)

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 <- readxl::read_xlsx("../data/intermediate/density.xlsx")
library(choroplethrAdmin1)
library(choroplethr)
admin1_choropleth("tunisia", density, num_colors = 7, title = "Population Density", legend = "Populatio
  ggtitle("Tunisia's Population Density by Governorate",
          subtitle = "Year: 2014") +
  labs(caption = "Source: L'Institut National de la Statistique (INS)") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```



Source: L'Institut National de la Statistique (INS)

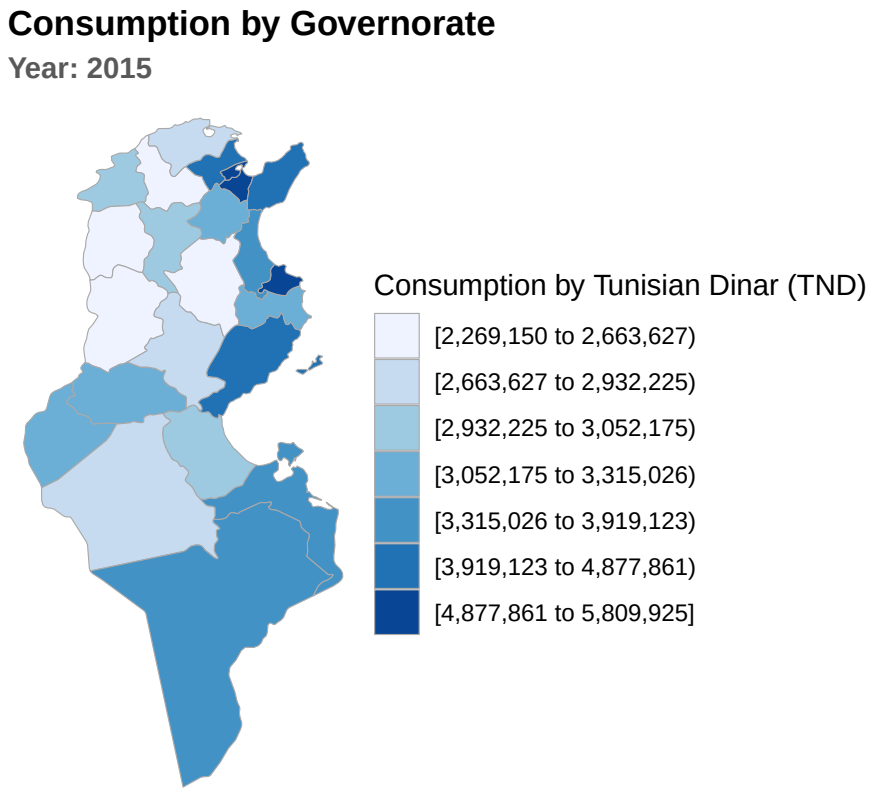
Comments: - coastal - point out tunis - Note: (1/24) Gov Ariana not included in the ChoroplethrAdmin 1 library

(4.b) Consumption by Governorates

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

```
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", "No

#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
```

```
#as.numeric(as.character(exitDF$immigrant))

df = data.frame(region=tidyConsump15$Governorates, value=as.numeric(as.character(tidyConsump15$Consumpt

admin1_region_choropleth(df, legend = "Consumption by Tunisian Dinar (TND)") +
  ggtitle("Consumption by Governorate",
          subtitle = "Year: 2015") +
  labs(caption = "Source: L'Institut National de la Statistique (INS)") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```

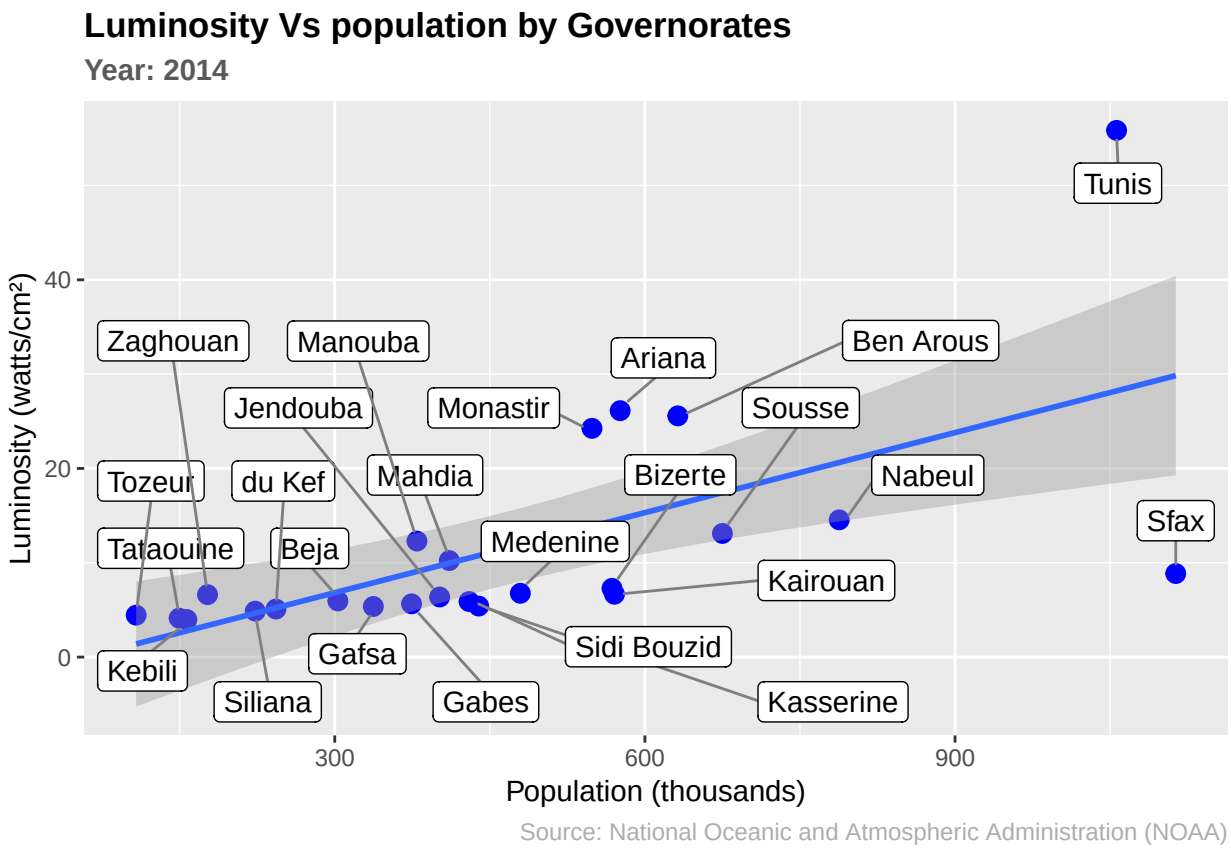


Comments: - definie consumption - can include excat categories that made up total - per capital - first comma from the right is delignating cents - show top 5 - focus on sfax

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

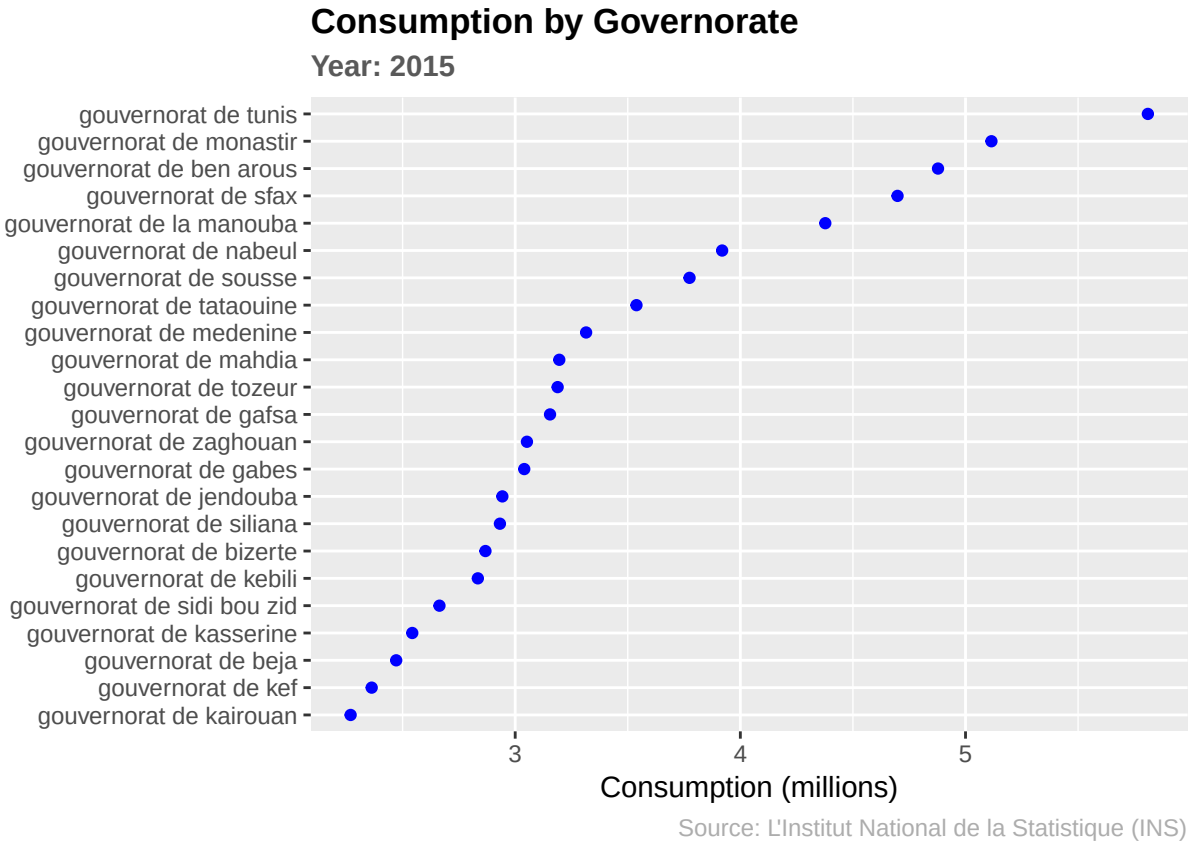
```
dataset <- readxl::read_xlsx("../data/intermediate/dataset.xlsx")
library(ggrepel)

dataset$population <- dataset$population / 1000

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

```
          subtitle = "Year: 2014") +
  labs(x= "Population (thousands)", y="Luminosity (watts/cm\u00b2)", caption = "Source: National Oceani
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```



Comments: - sfax outlier - general comment: eg. as population goes up luminsity goes up - "transition comment": Looking at consumption in descending order

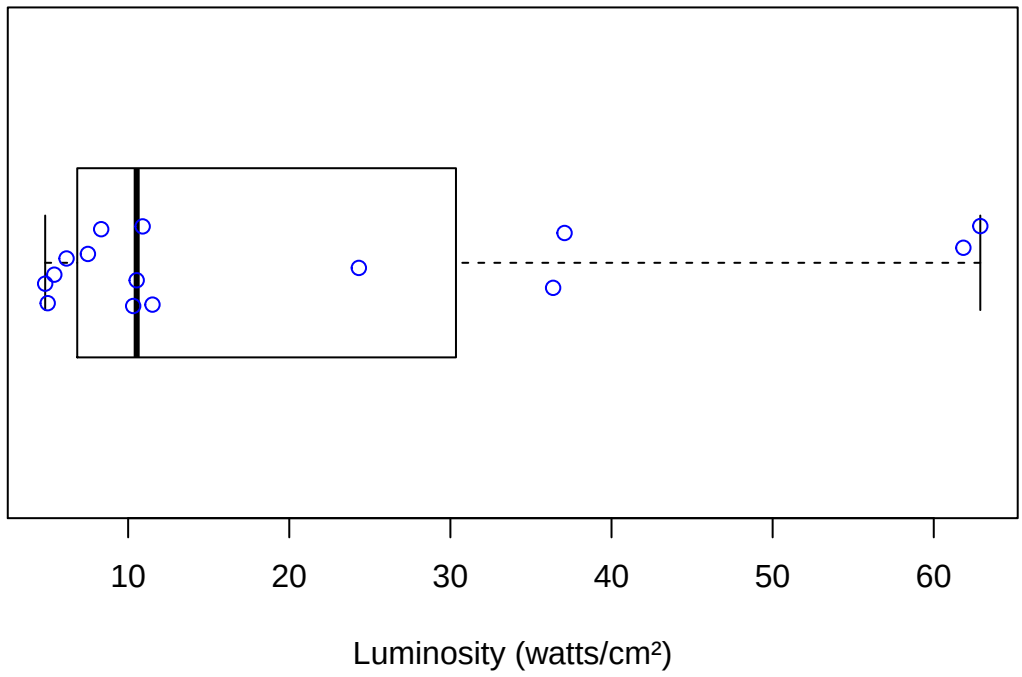
```
tidyConsump15$Consumption <- as.numeric(tidyConsump15$Consumption)/ 1000000
g <- ggplot(tidyConsump15, aes(x = Consumption, y = fct_reorder(Governorates, Consumption))) +
    geom_point(color = "blue") + ylab("") +
  ggtitle("Consumption by Governorate",
          subtitle = "Year: 2015") +
  labs(x = "Consumption (millions)", caption = "Source: L'Institut National de la Statistique (INS)") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
g
```



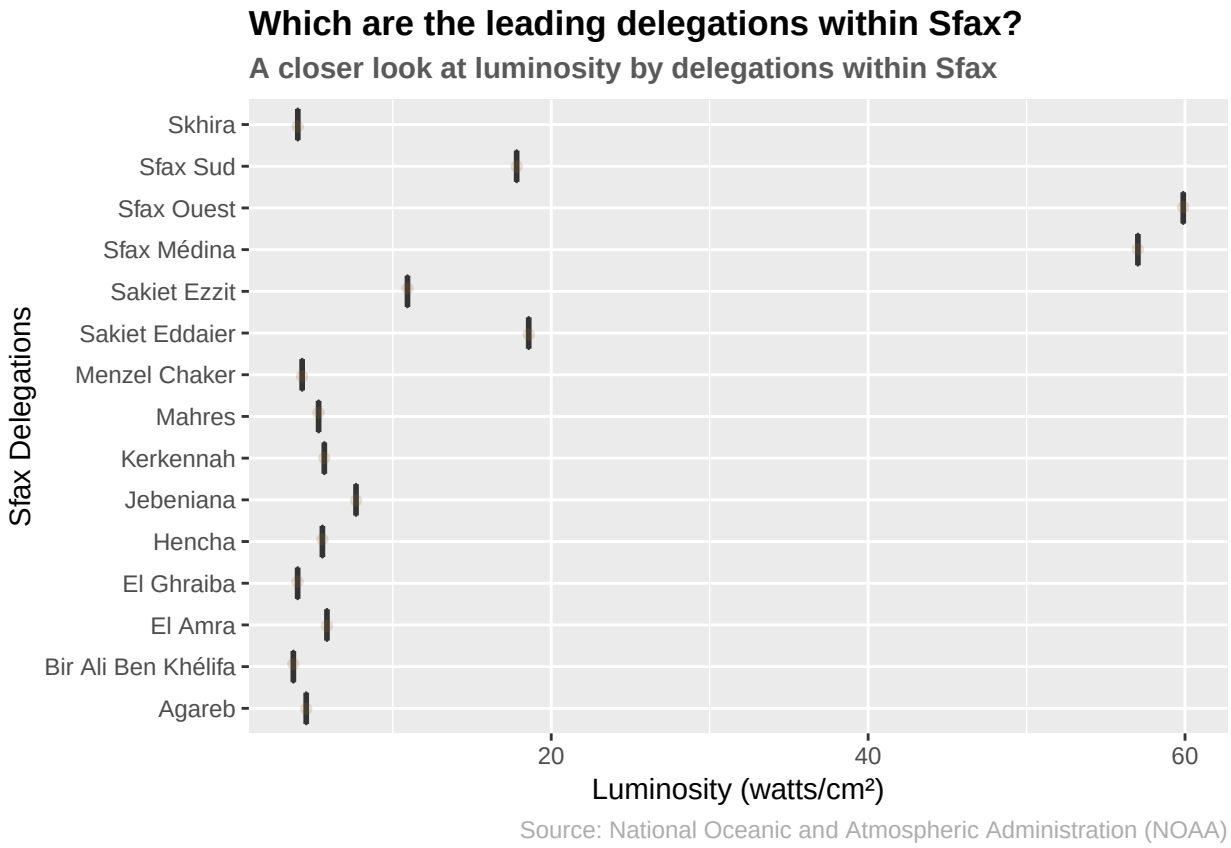
Comments: - note sfax is part of top 5 but is a clear outlier in the lowess plot

(4.d) A closer at outliers: Sfax

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

del_sfax <- filter(lum_del,lum_del$NAME_1 == "Sfax")
boxplot(del_sfax$`2013_mean`, horizontal = TRUE, main = "Luminosity of Sfax Delegations (Year: 2013)", x
stripchart(del_sfax$`2013_mean`, col = "blue", pch = 21, add = TRUE, method = "jitter")
```




```
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("Which are the leading delegations within Sfax?",
          subtitle = "A closer look at luminosity by delegations within Sfax") +
  labs(x= "Sfax Delegations", y="Luminosity (watts/cm\u00b2)", caption = "Source: National Oceanic and A
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```



Comments: - Clear unbalance between delegations within the Sfax governorates. The outliers are: xxx and YYY.

What might explain such variation in luminosity? - a little more

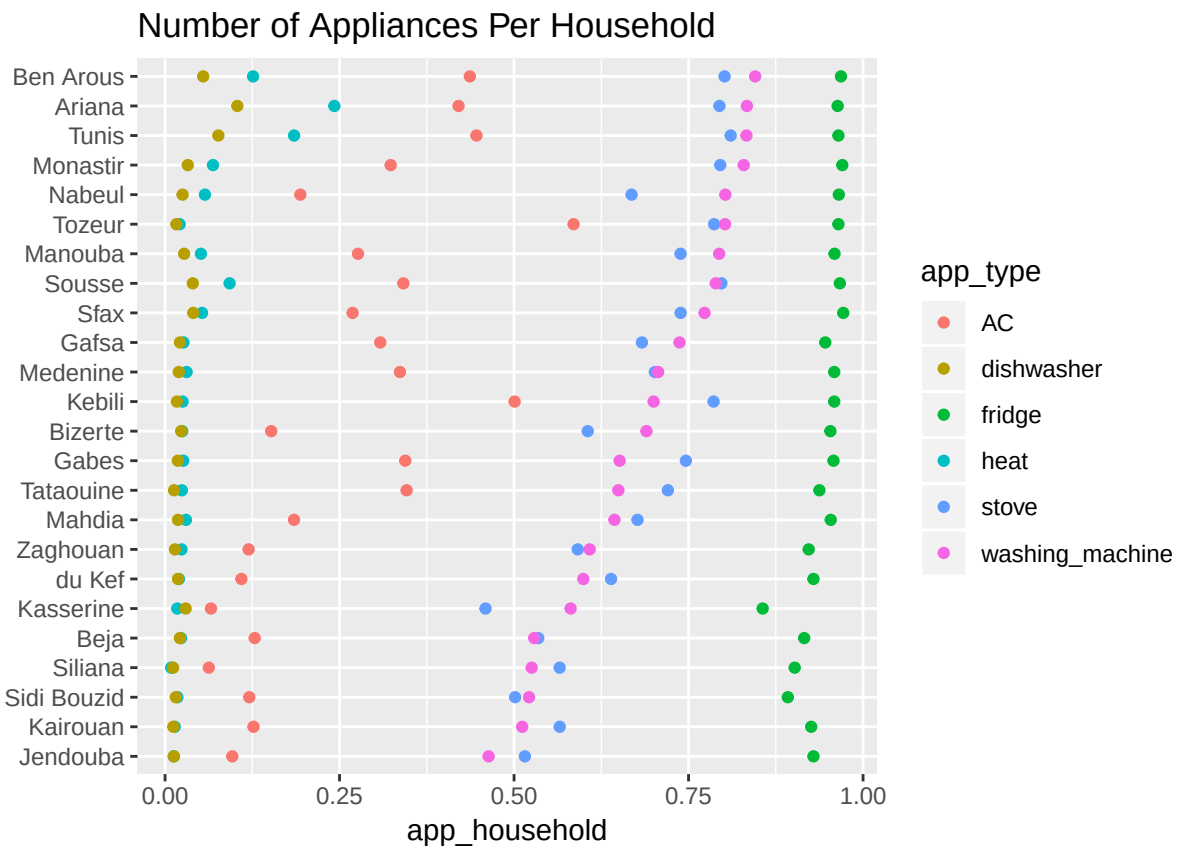
(4.e) Exploring potential covariates of luminosity

```
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
```

```r
appliances_total <- appliances %>%
  filter(Measure == "Total") %>%
  select(-Measure)

#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 ", ""))
```

```r
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