Nightlight Satellite Imagery as a Predictor for Economic Activity. Use Case: Tunisia.

Jesse P. Cahill
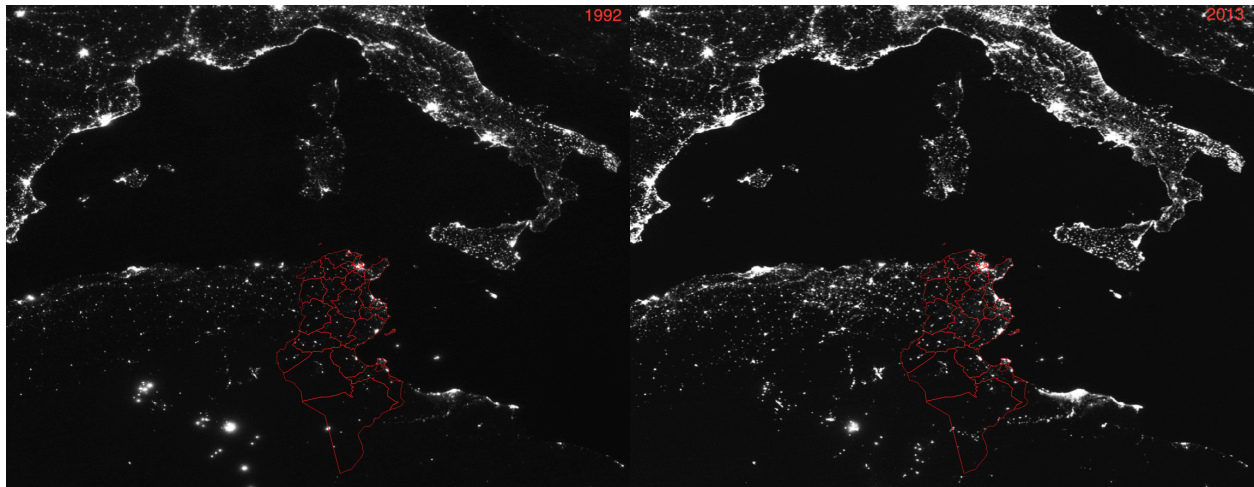*Mariem Ayadi*
Namson Ngo-Le

10 December 2018

(1) Introduction

```
img1992 <-  rasterGrob(as.raster(readPNG("../data/final/wYears/map_1992.png")), interpolate = FALSE)
img2013 <-  rasterGrob(as.raster(readPNG("../data/final/wYears/map_2013.png")), interpolate = FALSE)
grid.arrange(img1992, img2013, ncol = 2)
```


Visit the following link to view a year by year animation of nightlight imagery from 1992 to 2013:

https://github.com/mariemayadi/data/blob/master/luminosityOverTimeTunisia.gif

Why we chose this topic

We are studying how effective luminosity is as a predictor of economic activity in Tunisia. The primary reason for choosing this topic is to attempt to take a first step at filing a substantial gap in research. Many countries today, especially from the developing world, rely primarily on sporadic survey data such as the one conducted by the World Bank to produce their research. These surveys are not only sparse (they are rarely conducted on a yearly basis) but also expensive and at times not very granular. Another issue is the reliability of survey data. Depending on which organisation is responsible to produce the data, the output might include inaccuracies. Our proposed alternative, which we document in detail in our report, is to convert NASA's Nightlight Satellite Imagery into a luminosity value. The data is (1) publicly available, (2) has been measured on a yearly basis (as well as on a more granular time scale in some cases) for at least 20 years, and (3) covers almost the entire inhabitable earth.

Questions we are interested in studying

Luminosity data has been observed to be an effective proxy for economic activity, especially in developing countries with poorer statistical systems (Chen 2010). In particular, statistical measures of nighttime lights are correlated with measured income and GDP per capita (Pinkovskiy 2014). In addition, measures of luminosity can also tell us about the access to electricity and the level of public goods provision in a given region (Michalopoulos 2012).

Given that this data can be very objectively and consistently measured, has been gathered over such a long time-span, and can be measured at a very high level of geographical fineness (Henderson 2009), it is likely to be conducive to enhancing our understanding of economic activity in a given country. As Elvidge et al. put it, "Nighttime lights provide a useful proxy for development and have great potential for recording humanity's presence on the earth's surface and for measuring important variables such as annual growth rates for development" (Elvidge 2007). Finally, luminosity data has many useful data visualization purposes; maps of nighttime lights are both nice to look at and insightful.

Practical definitions:

Tunisia is located in Northern Africa. Its capital is Tunis. The country is divided into 7 broad regions: Greater Tunis area, North East, North West, Center East, Center West, South East and South West.

Each region is composed a number of governorates; in total, Tunisia has 24 Governorates. (A "Governorate" is the equivalent to a "State" in the United States of America or a "Province" in Canada. Note: The ChoroplethrAdmin1 package provides a breakdown of Tunisia by governorates.)

Furthermore, each governorate is comprised of "Delegations". In total, Tunisia has 264 delegations.

```
imgLoc <-  rasterGrob(as.raster(readPNG("../img/Tunisia-location-globe.png")), interpolate = FALSE)
imgGov <-  rasterGrob(as.raster(readPNG("../img/Tunisia-governorates.png")), interpolate = FALSE)
grid.arrange(imgLoc, imgGov, ncol = 2)
```

Team contribution

The work was evenly split between our three team members: Jesse P Cahill, Mariem Ayadi and Namson Ngo-Le. Initially, two of us teamed up to familiarize ourselves with using QGis to ingest and process the NOAA satellite imagery. Another team member focused on reading relevant research papers and extracting data from the Tunisian National Institute of Statistics (INS). At a later stage, each member was assigned a third of all of the collected input data and was tasked to create insightful exploratory graphs that would help answer our research questions. As a team, we then were able to modify, integrate and generate a report based on the created visuals.

(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

Satellite Data

Satellite data was collected from the National Oceanic and Atmospheric Administration (NOAA) Version 4 DMSP-OLS Nighttime Lights Time Series. The data is located at https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html The images are cloud-free composites made using all the available archived DMSP-OLS smooth resolution data for calendar years 1992 through 2013. In some years two separate satellites were collecting data, and in those years we defaulted to the images from the older satellite. We downloaded the compressed versions of these files as geotiffs and loaded them into QGIS as 13 individual raster layers. The images are originally very dark so not much data is apparent on the standard brightness scale of 0 as black and 255 as maximum brightness. To rectify this the images were rescaled from 0 to 63 by the provider prior to downloading, thus making differences in luminosity more visually apparent.

Our goal of examining differences in data by Tunisian governorate required us to obtain summary statistics for each individual governorate from every raster image. This was obtained by imposing a vector layer of polygons making up each governorate on top of the raster satellite images in the form of a shapefile, which was loaded into QGIS. The shapefile was sourced from GADM, the Database of Global Administrative Areas. We used shapefiles for both governorates (analogous to a U.S. State) and delegations (analogous to a U.S. county).

We then used the QGIS zonal statistics tool to gather aggregated luminosity data for each of the 24 governorates and the 264 delegations for each year's satellite images. The zonal statistics tool works by treating each pixel value within each geographic unit designated from the shapefile as a separate datum and each geographic unit as a separate dataset. Summary statistics are then computed based on those data values. We are most interested in the following statistics:

- **Count**: Total count of the number of pixels within the geographic unit. Because each pixel has a set area, this can be used as a proxy for geographic unit area.
- **Sum**: Total sum of all pixel values within the geographic unit.
- **Mean**: Sample mean of all pixel values within the geographic unit. This is our most used summary statistic and our primary means of comparing the luminosity between governorates, as it is not dependent on area.
- **Variance**: Sample variance of all pixel values within the geographic unit. We believe this can be used to identify governorates that are small and very uniformly lit (which should have lower variance) from governorates that have some highly lit areas and other very dark areas, even though the two may have similar mean luminosity.

Note: Relevant QGis screenshots have been added to the Appendix.

Economic Data

Most economic statistics were collected from the Tunisia Data Portal at http://dataportal.ins.tn/en/. This website is run by the L'Institut National de la Statistique (INS) of Tunisia. We were primarily interested in governorate-level data as opposed to national-level accounts, as we wanted to study within-country variation in luminosity and economic activity. We chose to focus on Tunisian economic data from 2014 because studies have indicated that luminosity is a better indicator for cross-sectional as opposed to panel/time-series data (Chen 2010). The data came from the Tunisia's government 2014 Census.

We were able to download csv files of statistical data for variables such as households by possession of leisure resources and Information and Communication Technologies (ICT's), access to public services, and workers in particular industries (e.g. agriculture, manufacturing, mining). We then selected a few indicators that we believed were relevant to the comparison of luminosity to economic variables. Finally, we did some minor formatting changes of the csv files such as renaming variables and sorting rows and columns before exporting them into R for visualization and analysis.

We also collected some national-level time-series economic data for Tunisia from the IMF World Economic Outlook Database. These variables were total GDP, GDP per capita, GDP growth, inflation, unemployment rates, and government debt from 1992 until 2013 and were used in comparison to national-level measures of luminosity in our project.This data is mostly self-explanatory and easy to access at this link: https://www.imf.org/external/pubs/ft/weo/2018/01/weodata/weorept. aspx? pr.x=96&pr.y=0&sy=1980&ey=2023&scsm=1&ssd=1&sort=country&ds=.&br=1&c=744 &s=NGDP_RPCH,PPPGDP,PPPPC,PCPIPCH,LUR,GGXWDG_NGDP&grp=0&a=

(3) Analysis of data quality

Satellite Data

There are some concerns in general using longitudinal cloud-free composite images for luminosity data. Additional variance can be introduced because these images are created from many images taken over the year to avoid luminosity being obscured by cloud coverage. Additionally, not all photos are taken using the same satellite. Six separate satellites are used over the 21 years that make up our dataset, and our literature search indicated that there can often be significant differences in the values observed by different satellites. This can be observed in the time-series graph of average luminosity across Tunisia, where we often see sharp jumps in years when a new satellite is introduced.

The spatial resolution of raster satellite images is also slightly lower than could be desired. When viewing graphs in QGIS, occasionally the Tunisia governorate shapefile wouldn't match up perfectly with the borders of the country, and would move between years. With our limited GIS software knowledge we weren't able to address this issue, and although minor it is possible there could have been a small effect on the data.

Luckily, due to the nature of the data processing we had no NA's in satellite data, as every pixel is assigned a value without fail.

Economic Data

Because we had to download individual files for each topic of the census, our economic data is a bit fragmented between different files. For example, there are separate spreadsheets for 'households by possession of ICT's', 'households by possession of leisure resources' and 'households by households by sources of supply of drinking water". Nevertheless the economic data appeared to be complete and accurate for the indicators that were provided. There was no missing data for any variable that we used for our project.

In addition, there were no circumstances in which we looked through the data and thought that there was something off about the numbers. Rich areas had more leisure resources, governorates with more population had higher total number of households, and so forth.

The biggest challenge with the economic data was that the definitions of the variables were often vaguely defined, which made us unsure of how to incorporate them into our analysis. For example, one variable as defined as 'Number of household owning a Parable', which was NOT the number of households that had a moral lesson to tell us, but rather the number of households that owned a dish antenna. We suspected that some of these shoddy definitions came down to meaning being lost in translation, given that Tunisia is an Arabic and French-speaking country. Overall our approach was simply to prioritize those variables that (1) had to do with luminosity (2) were more clearly defined to incorporate into our analysis.

(4) Main analysis (Exploratory Data Analysis)

(4.a) A First Look at Tunisia

```
density <- readxl::read_xlsx("../data/intermediate/density.xlsx")
library(choroplethrAdmin1)
library(choroplethr)
admin1_choropleth("tunisia", density, num_colors = 7,
                title = "Population Density",
                legend = "Population Density (km\u00b2)") +
  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"))
```

Tunisia's Population Density by Governorate
Year: 2014
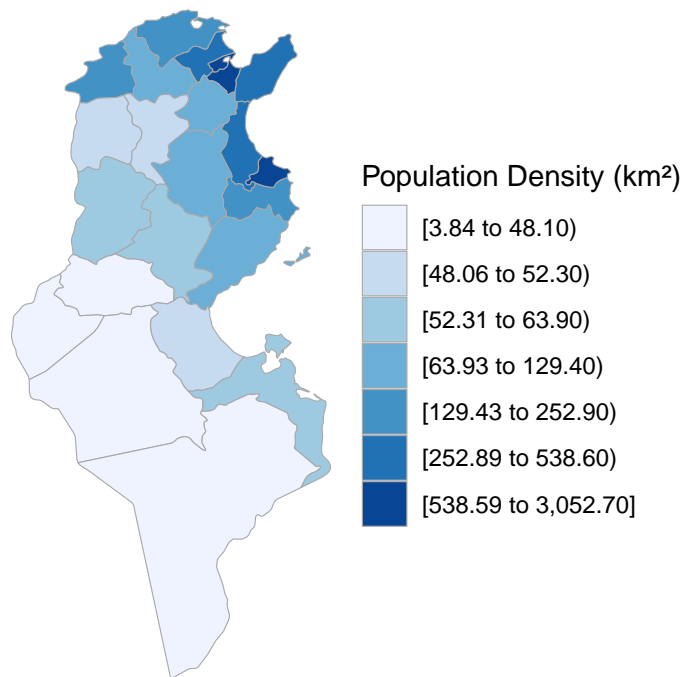
Source: L'Institut National de la Statistique (INS)

Tunisia's population is heavily concentrated on the coast and near the capital city of Tunis in particular,

while the larger inland regions are very sparsely populated. It is likely that indictors of economic activity as well as measures of luminosity will be heavily concentrated around Tunis and the rest of the coast.

Note that the governorates of Ariana and Manouba north of Tunis are combined in the choroplethrAdmin1 package, so all choroplethr plots of Tunisia on the governorate level will show Ariana and Manouba as a combined governorate.

(4.b) Population breakdown by Industry and Gender

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

occupation <- read_tun_data(here("data/raw",
"Distribution of the occupied population aged 15 years and over by the activity sector 11_18_2018 02_45_
  mutate(Region = str_replace(Region, "Governorate of ", "")) %>% # Standardize governorate names
  filter(Region != "Tunisia")


occupation <- occupation %>%
  mutate_at(vars(contains("_")), function(x) as.numeric(x) / house_pop$over_15_population) %>%
  select(-ends_with("Total"))

occupation <- occupation %>%
  gather(contains("_"), key="key", value="value") %>%
  separate(key, into=c("sector", "gender"), sep="_") %>%
  spread(sector, value)
```

```r
library(GGally)
# add y label - percent of total pop over 15 employed
occ_names <- c("Region", "gender", "Agr./Fishing", "Pub. Works", "Education/Health/Admin",
"Manufacturing", "Mines/Energy",
"Other Serv.", "Trade", "Transport", "Undeclares")
names(occupation) <- occ_names
ggparcoord(occupation, columns = 3:11, alphaLines = .5,
           scale = "uniminmax", splineFactor = 10, groupColumn = 2) +
  geom_vline(xintercept = 2:9, color = "lightblue") +
    ggtitle("Gender Disparity by Industry",
          subtitle = "Year: 2014") +
  labs(x = "Industry", y= "Percent employed",
  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"))
```

6

Gender Disparity by Industry
Year: 2014

Source: L'Institut National de la Statistique (INS)
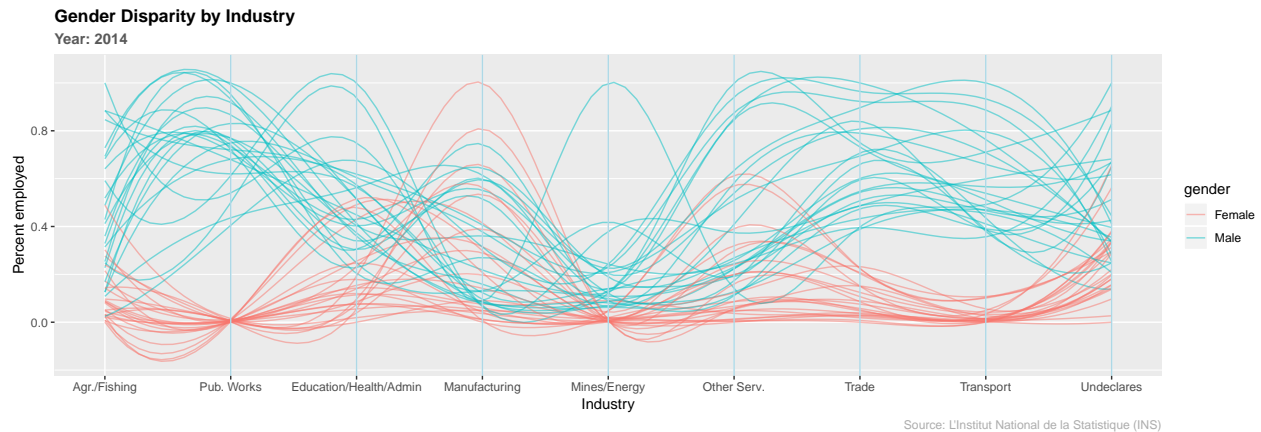
This plot shows one line per governorate per gender for the year 2014, with the Y value representing the fraction of total persons fifteen years or older that are employed in that industry. It allows us to observe disparities between employment sector trends between men and women, as well as which employment sectors tend to to be positively and negatively associated with one another. We see that in general, more men are employed than women despite the gender distribution of population being relatively equal. Some industries (like public works and energy and mining) show nearly no participation by women across the board. Others, like manufacturing, show a wide array of values for both genders. The highest peak of manufacturing by a decent margin is actually for women. It's difficult to discern any significant correlations between industries from this plot.

(4.c) Consumption by Governorates

```
consumption2015 <-
readxl::read_xlsx("../data/intermediate/Enquête Consommation 2010 12_08_2018 10_49_06.xlsx")

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,]

#removing Greater regions
tidyConsump15 <- filter(tidyConsump15, !Governorates %in%
    c("Great Tunis", "Governorate of Ariana", "North East",
      "North West","Par grandes régions\\Centre Est",
      "Par grandes régions\\Centre Ouest","Par grandes régions\\Sud Est",
      "Par grandes régions\\Sud Ouest"))

#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 manouba", "gouvernorat de nabeul",
    "gouvernorat de zaghouan", "gouvernorat de bizerte",
    "gouvernorat de beja","gouvernorat de jendouba","gouvernorat de kef","gouvernorat de siliana", "gou
```
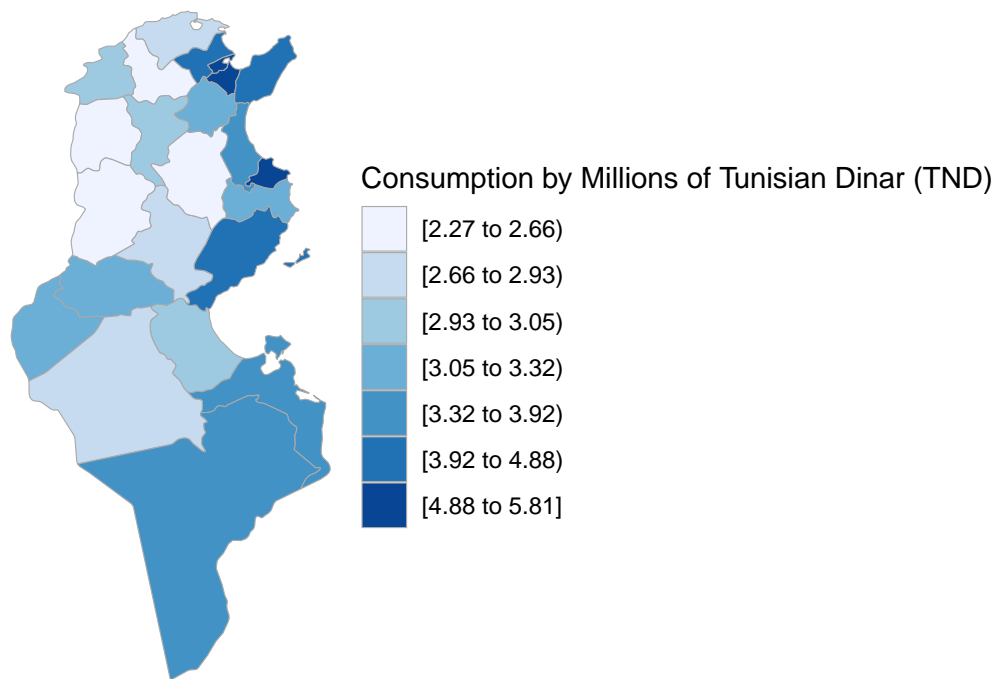
```
    "gouvernorat de monastir", "gouvernorat de mahdia",
    "gouvernorat de sfax", "gouvernorat de kairouan",
    "gouvernorat de kasserine", "gouvernorat de sidi bou zid",
    "gouvernorat de gabes", "gouvernorat de medenine",
    "gouvernorat de tataouine", "gouvernorat de gafsa",
    "gouvernorat de tozeur","gouvernorat de kebili")


df = data.frame(region=tidyConsump15$Governorates, value=
as.numeric(as.character(tidyConsump15$Consumption)) / 1000000)

admin1_region_choropleth(df,
  legend = "Consumption by Millions of 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"))
```

Consumption by Governorate

Year: 2015

Consumption by Millions of Tunisian Dinar (TND)

- [2.27 to 2.66)
- [2.66 to 2.93)
- [2.93 to 3.05)
- [3.05 to 3.32)
- [3.32 to 3.92)
- [3.92 to 4.88)
- [4.88 to 5.81]

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

The variable consumption here is defined as the average consumption per capita of each governorate in 2015. Total consumption is a sum of the consumption of products and services such as alcohol, furniture, cultural activities, restaurants, and transportation for each governorate. The units are in the Tunisian currency dinar, and the first comma from the right is a separator for the amount of millimes and dinars (which are roughly analogous to American cents and dollars, with the difference being that 1000 millimes = 1 dinar). Thus for example the highest measure of consumption per capita in the data set is Tunis with 5,809 dinars and 925

millimes.

While the wealthy coastal governorates have the highest levels of consumption per capita, there are some subtle differences from the population density choroplethr plot. Consumption is lowest in some of the inland northwest governorates, while the southern, sparsely populated governorates have relatively high levels of consumption per capita compared to the rest of the country.

Overall there appears to be a lot more economic activity in the densely populated coastal regions of Tunisia as well as disparities in industry participation by gender. Incorporating our luminosity data into our analysis should prove useful in gaining better insights into some of these subnational-level variations in economic data.

**(4.d) Luminosity trends across Tunisia**

```r
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(label = governorate),box.padding   = 0.35,
                    point.padding = 0.5,
                    segment.color = 'grey50') +
  ggtitle("Luminosity Vs population by Governorates",
          subtitle = "Year: 2014") +
  labs(x= "Population (thousands)", y="Luminosity (watts/cm\u00b2)",
       caption = "Source: National Oceanic and Atmospheric Administration (NOAA)") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```

Luminosity Vs population by Governorates

Year: 2014
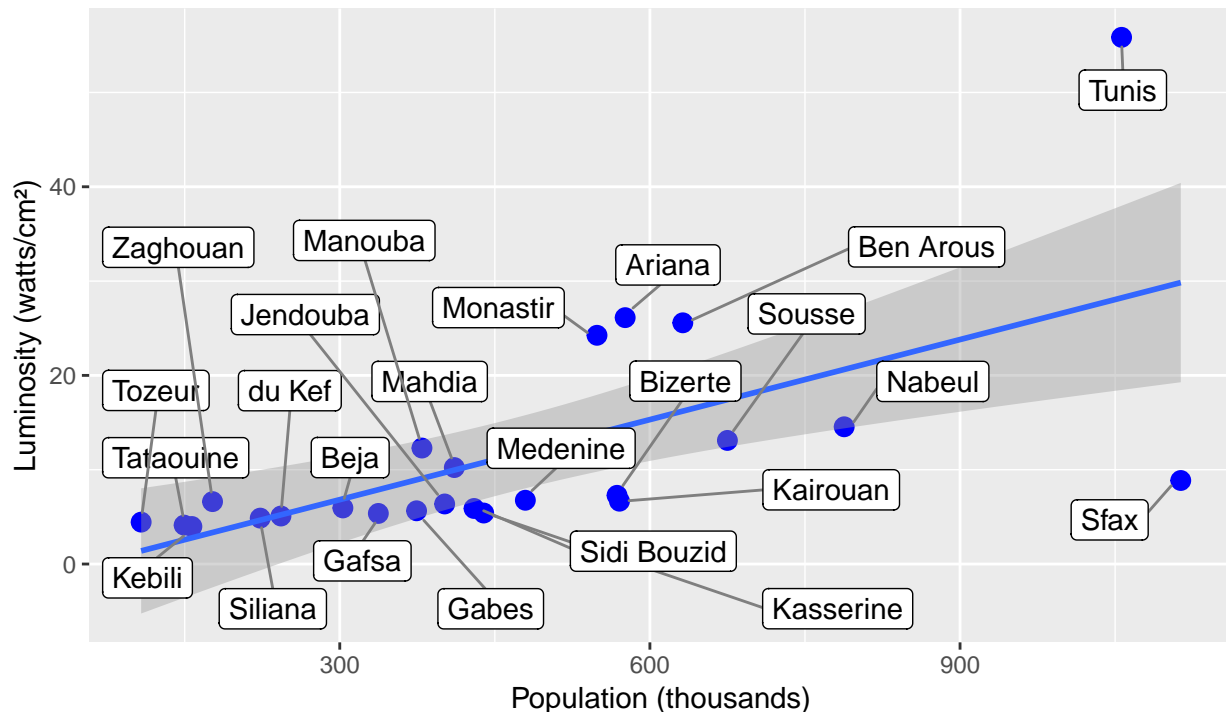
Source: National Oceanic and Atmospheric Administration (NOAA)

This scatter plot of mean luminosity (in watts/cm2) against population (in thousands) by governorate shows some interesting trends, especially for the two high population governorates of Tunis and Sfax to the right of the graph. Overall with the loess smoother, we can see that as expected there is a general upward trend in that as population increases, luminosity also increases. However, Tunis is significantly above the trendline, suggesting that it has a extraordinarily high level of luminosity compared to its population (which is 1,056,247 million to be exact). This further reinforces the point that commercial activity in Tunisia is heavily concentrated in Tunis.

In contrast, Sfax (population 1,113,496) is sharply below the trendline. While Sfax is even more populous than Tunis, it has a significantly lower level of mean luminosity relative to population than the rest of Tunisia. This is a rather bizarre finding that merits further exploration. We decided to take a look at consumption per governorate to see if it could tell us anything about the outlier of Sfax.

```
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 (thousands)",
       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
```

Consumption by Governorate
Year: 2015
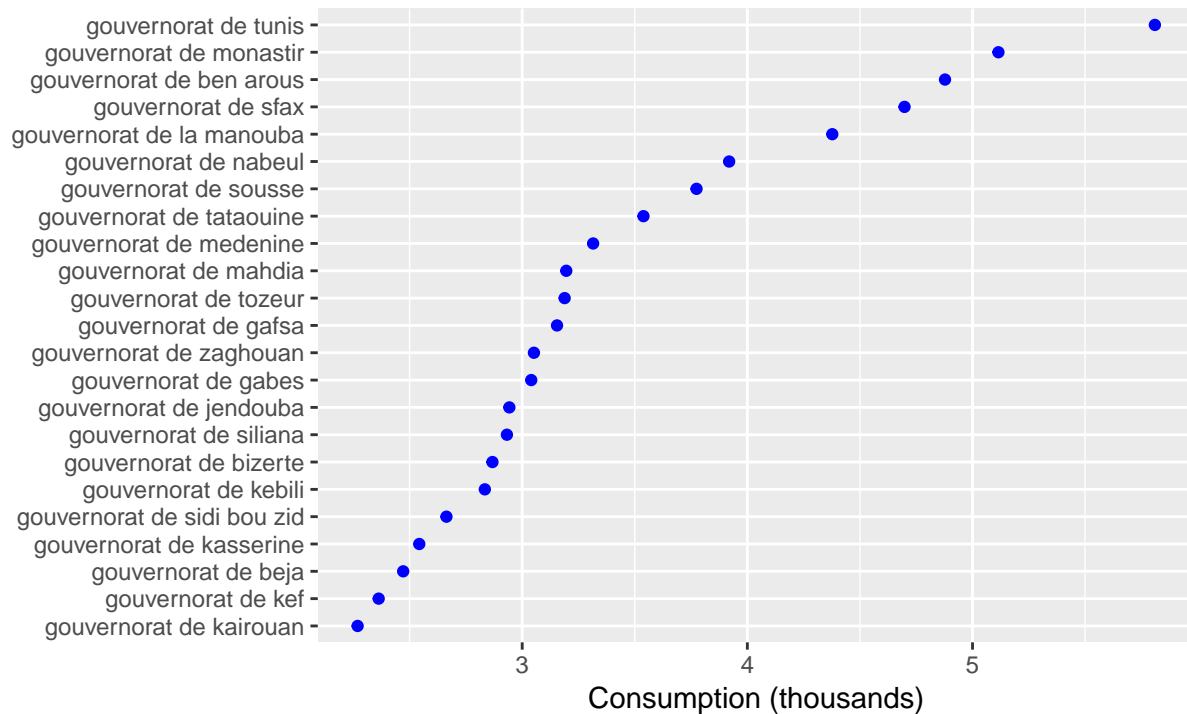
Source: L'Institut National de la Statistique (INS)

This is a dot plot of consumption per capita (see 4.c for a thorough definition of this variable) by governorate for the year 2015. Tunis has the highest consumption per capita in the country, while Monastir, Ben Arous, and Manouba, which are essentially three suburban regions close to Tunis geographically are also fairly close to it in terms of consumption. Sfax appears as having a high level of consumption per capita, despite having low luminosity as we have seen in the previous graph. Therefore even by per capita measures Sfax seems to have a high level of economic activity, yet low total luminosity. We take a closer look at luminosity in Sfax in the next section.

The rest of the governorates are fairly close to one another, suggesting that there are a few governorates in Tunisia that make up a a disproportionate share of the consumption in the country.

Note that we divided the consumption figure by a million, so the x-axis in this graph is consumption in thousands of Tunisian Dinars (TND).

(4.e) 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)", xlab = "Luminosity (watts/cm\u00b2)")
stripchart(del_sfax$`2013_mean`, col = "blue", pch = 21, add = TRUE, method = "jitter")
```

Luminosity of Sfax Delegations (Year: 2013)


Luminosity (watts/cm²)
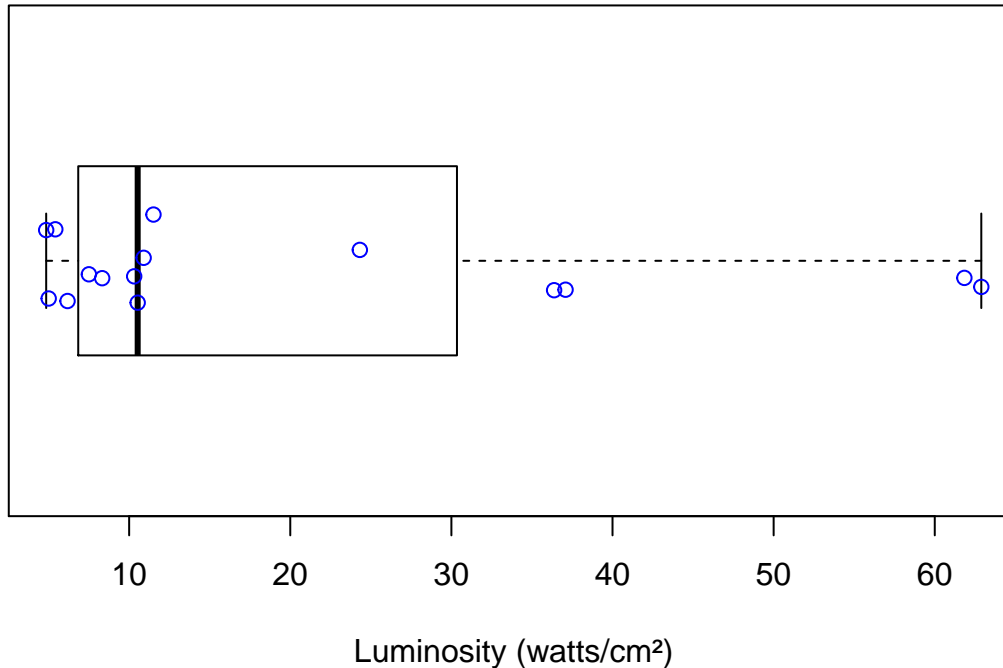
To further examine the unusually low value of mean luminosity for the governorate of Sfax, we decided to examine it at a higher level of spatial granularity. We grouped the delegation level data for all 16 delegations within Sfax to gain a better understanding of its low mean luminosity value. After examining the distribution with a boxplot and jittering the points to add visibility, it became clear that the distribution of mean delegation luminosity is largely clustered around a value of around 10/63, but has a very long tail going to the right with two delegations having values of nearly 63. Our hypothesis was that these two delegations represent the city of Sfax, which is a major port city with a large population. However, because these areas are very small and the governorate as a whole is large, the mean luminosity value is dragged much lower by these other more rural delegations. This was confirmed in the following plot which shows the values for each delegation within Sfax by name.

```
ggplot(del_sfax, aes(del_sfax$NAME_2, del_sfax$`2013_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 Atmospheric Administration (NOAA)") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```

Which are the leading delegations within Sfax?

A closer look at luminosity by delegations within Sfax
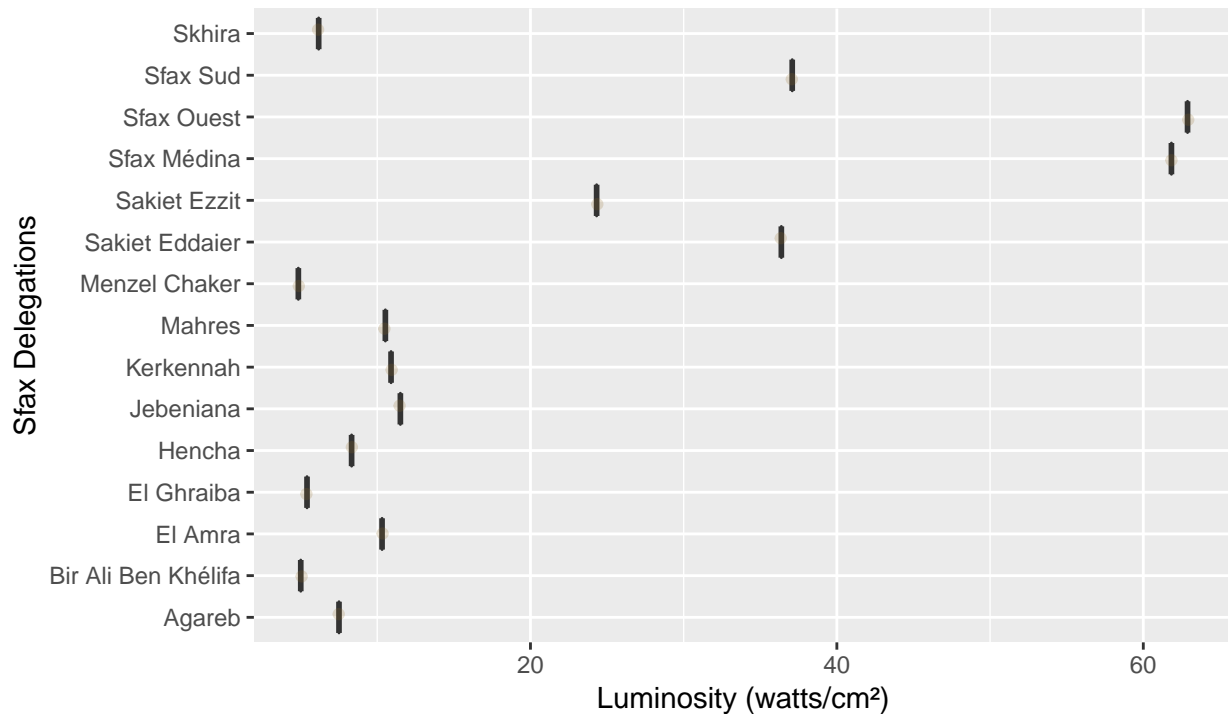
Source: National Oceanic and Atmospheric Administration (NOAA)

(4.f) Exploring potential covariates of luminosity

```r
# appliance data

appliances <- read_tun_data(here("data/raw",
 "Households by possession of electrical household equipements 11_18_2018 02_51_51.xlsx")) %>%
  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 %>%
```

```r
  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 ", ""))

appliances_tidy <- appliances_total %>%
  gather(-Region, -households, key="appliances_type", value="num_appliances") %>%
  mutate(appliances_household = num_appliances / households)

g_stacked <- ggplot(appliances_tidy, aes(y = appliances_household,
                               x = fct_reorder2(Region, appliances_type, -appliances_household),
                               fill = appliances_type)) +
    coord_flip() +
    geom_bar(stat="identity") + ylab("") +
    ggtitle("Appliances Resources per Household",
          subtitle = "Year: 2014") +
  labs(x = "Number of Appliances per Household", y= "Governorates",
       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_stacked
```
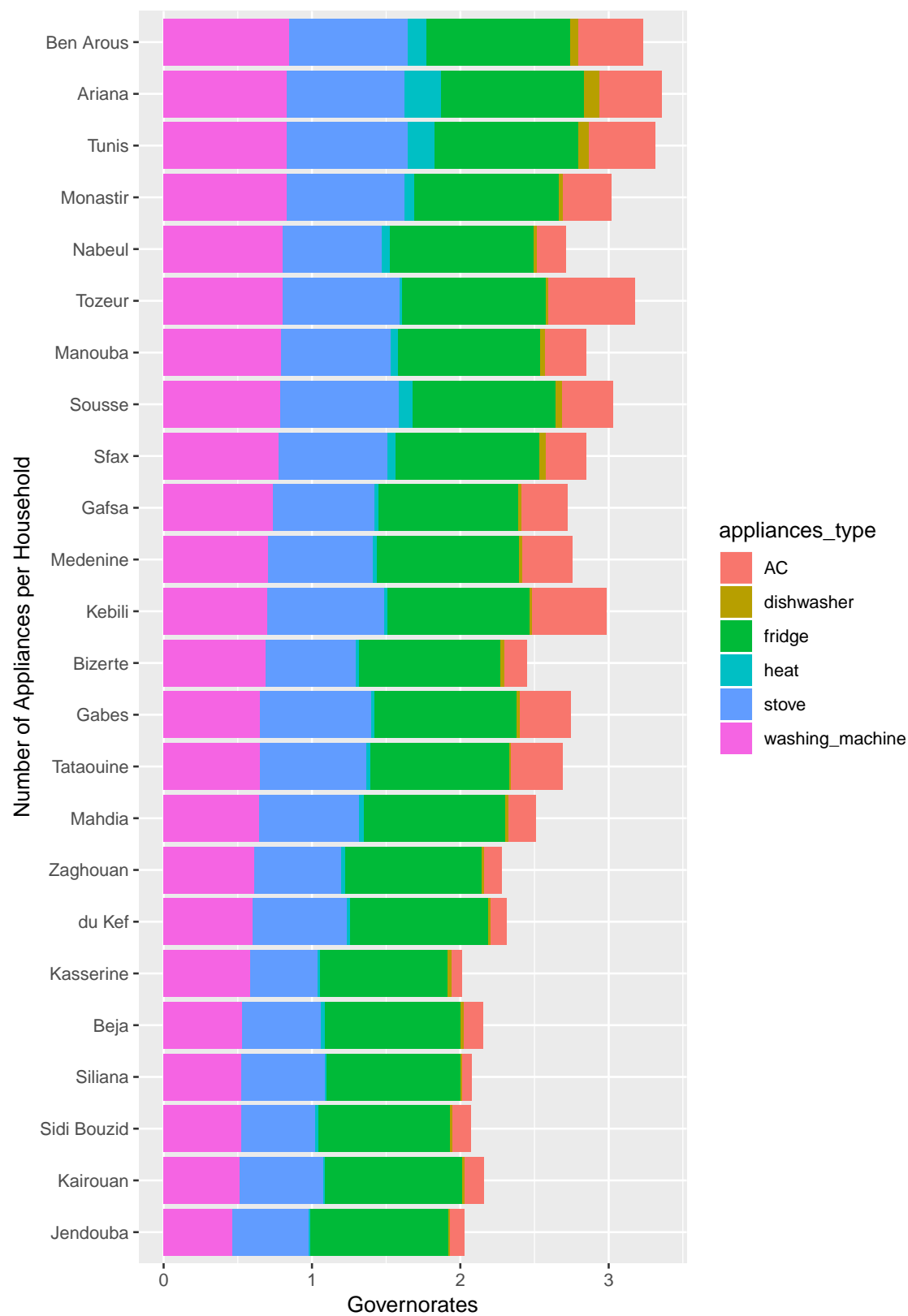
Appliances Resources per Household

Year: 2014

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

This plot compares the numbers of functional appliances per household by governorate. Appliances examined are air conditioners, dishwashers, refrigerators, heaters, stoves, and washing machines. Refrigerators are the most ubiquitous appliance, with the average per household being nearly 1 across the board. Next are stoves and washing machines which are still relatively common, but with some governorates having only around .5 per household. Dishwashers are extremely uncommon even in urban governorates like Ariana and Tunis. Heating is similar. Air conditioning shows a decent amount of variation between governorates and doesn't seem to be purely associated with urbanicity. In future work it would be interesting to look at how correlated number of air conditioners is with the average temperature of the governorate. A common theme among the total average number of appliances per household is that landlocked governorates (which also tend to have lower consumption) tend to have less resources.

```r
internet <- read_tun_data(here("data/raw",
                               "Households by possession of ICTs 11_18_2018 02_51_58.xlsx")) %>%
  mutate(Region = str_replace(Region, "Governorate of ", "")) %>% # Standardize governorate names
  filter(Region != "Tunisia") %>%
  select(-contains("households_"))

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

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

# we only want total
internet_total <- internet %>%
  filter(Measure == "Total") %>%
  select(-Measure)

#fix the names
internet_names = c("Region", "households", "computer", "landline", "internet", "mobile")
names(internet_total) <- internet_names

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

#graph it

internet_tidy <- internet_total %>%
  gather(-Region, -households, key="ict_type", value="num_ict") %>%
  mutate(ict_household = num_ict / households)

g_stacked <- ggplot(internet_tidy, aes(y = ict_household,
                                       x = fct_reorder2(Region, ict_type, -ict_household),
                                       fill = ict_type)) +
  coord_flip() +
  geom_bar(stat="identity") + ylab("") +
  ggtitle("Information and Communications Technology (ICT) Resources per Household",
          subtitle = "Year: 2014") +
  labs(x = "Number of Appliances per Household", y= "Governorates",
       caption = "Source: L'Institut National de la Statistique (INS)") +
  theme(plot.title = element_text(face = "bold")) +
```
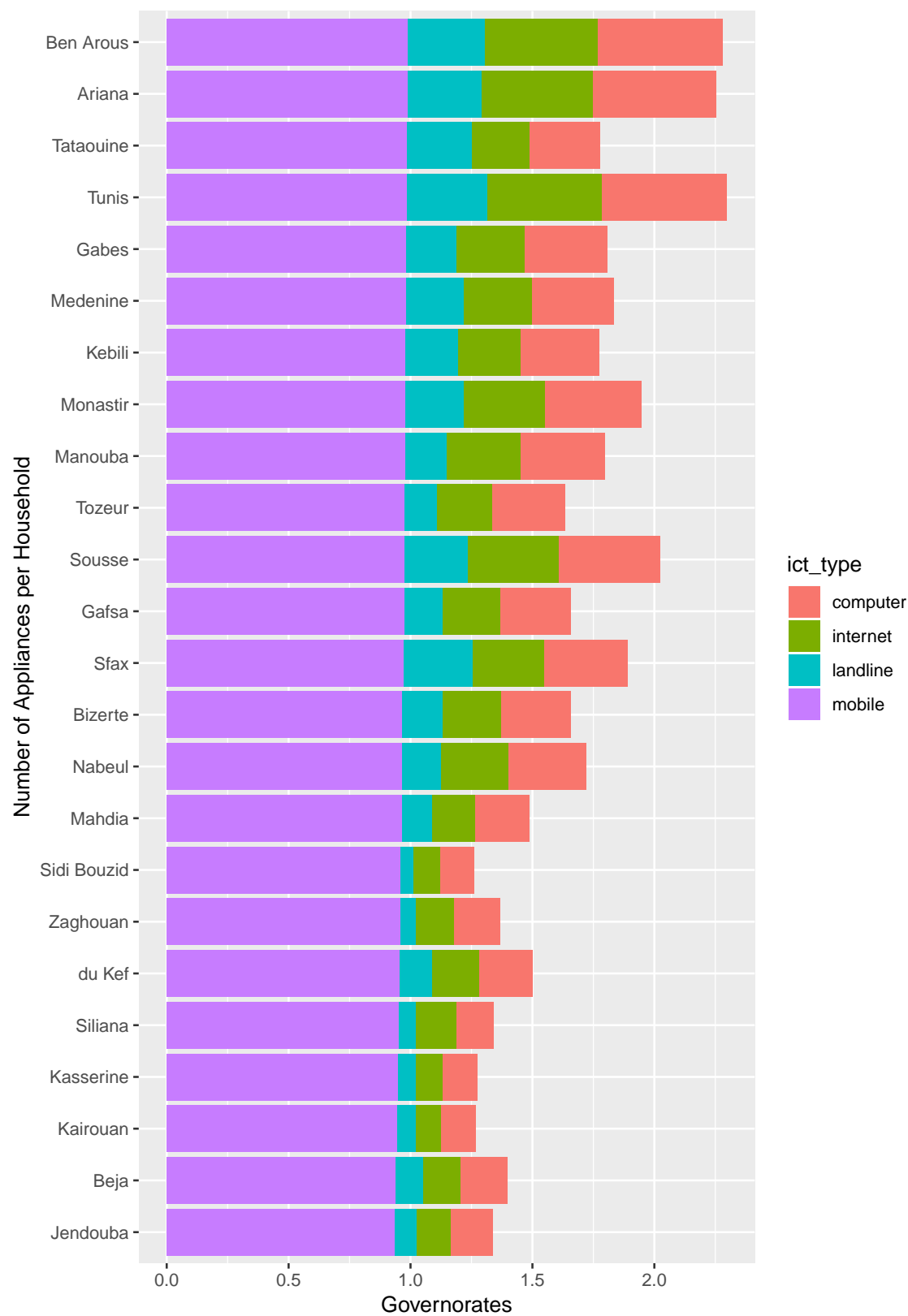
```
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))

g_stacked
```

Information and Communications Technology (ICT) Resources pe

Year: 2014

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

This plot compares the number of Information and Communication Technologies (ICT) resources per household by governorate. ICTs examined are mobile phones, landline phones, internet connections, and computers. We quickly see that mobile phones have a similar role in this plot to refrigerators in the previous one; nearly every household has at least one. Landlines vary more and more common in urban centers like Sfax, Tunis, Ariana and Ben Arous. There seems to be a correlation between landlines and internet connections, and the numbers often appear to be very similar. This may be because the same companies service phone landlines and internet connections, and bundle the two. Numbers for computers also appear to very similar to internet connections as well, even more so than landlines. We observe the same phenomenon here for total number of ICTs per household as in appliances, where we see less total ICTs in landlocked governorates (although Tataouine and Kebili do not quite follow these trends).

Leisure Resources

```r
leisure <- read_tun_data(here("data/raw",
                              "Households by possession of leisure resources 11_18_2018 02_51_43.xls
  mutate(Region = str_replace(Region, "Governorate of ", "")) %>% # Standardize governorate names
  filter(Region != "Tunisia") %>%
  select(-contains("households_"))

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

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

# we only want total
leisure_total <- leisure %>%
  filter(Measure == "Total") %>%
  select(-Measure)

#fix the names
leisure_names = c("Region", "households", "car", "library", "dish", "radio", "TV")
names(leisure_total) <- leisure_names

# cleaning up
leisure_total <- leisure_total %>% select("Region", "households", "car", "dish", "radio", "TV") %>%
  mutate_at(.vars = vars(-Region), funs(as.numeric)) %>%
  mutate(Region = str_replace(Region, "Governorate of ", ""))

leisure_tidy <- leisure_total %>%
  gather(-Region, -households, key="leisure_type", value="num_leisure") %>%
  mutate(leisure_household = num_leisure / households)

g_stacked <- ggplot(leisure_tidy, aes(y = leisure_household,
                                      x = fct_reorder2(Region, leisure_type, -leisure_household),
                                      fill = leisure_type)) +
    coord_flip() +
    geom_bar(stat="identity") + ylab("") +
    ggtitle("Leisure Resources per Household",
            subtitle = "Year: 2014") +
```
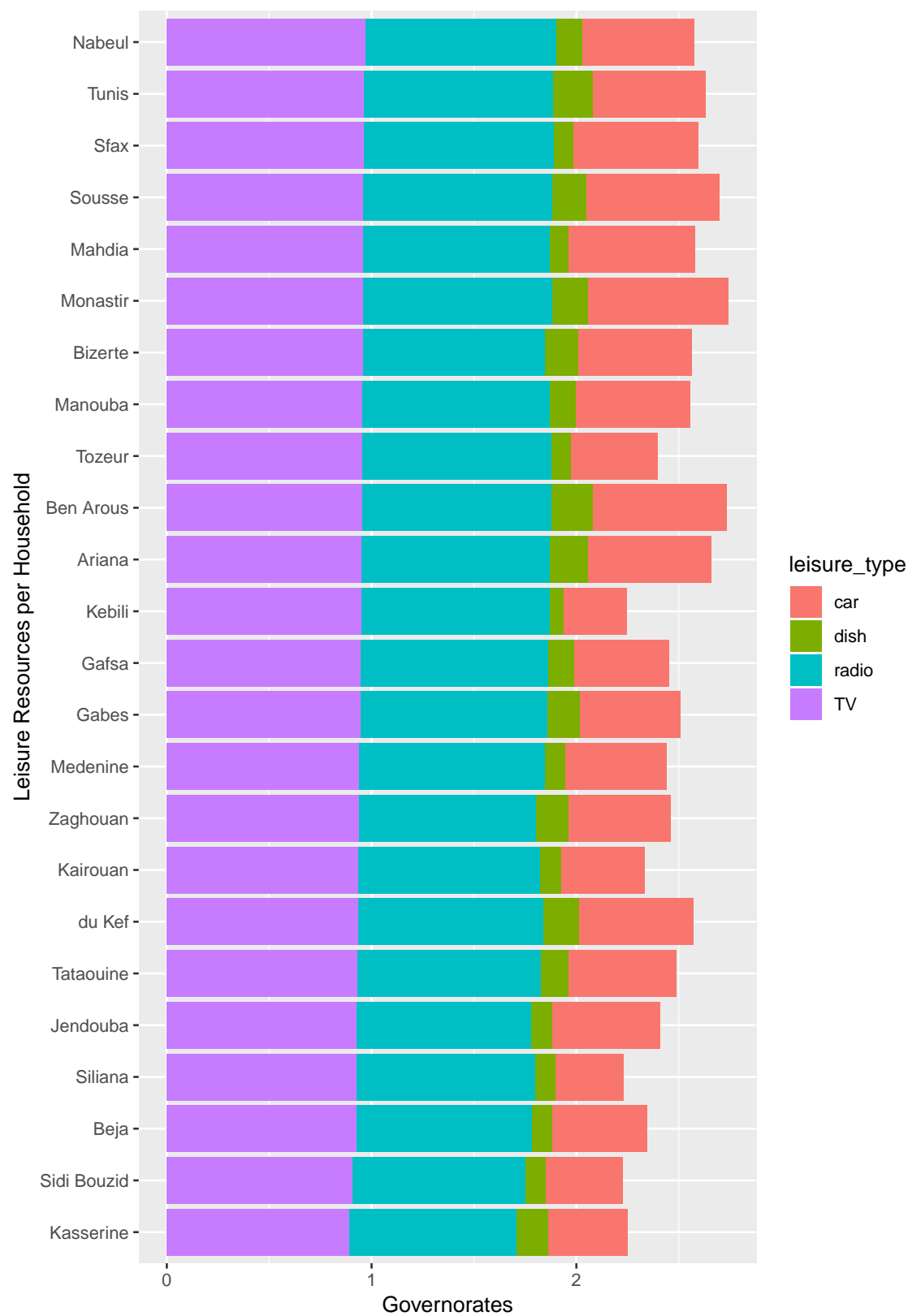
```r
  labs(x = "Leisure Resources per Household", y= "Governorates",
       caption = "Source: L'Institut National de la Statistique (INS)") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))

g_stacked
```

Leisure Resources per Household

Year: 2014

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

21

This plot compares the number of leisure resources per household by governorate. Leisure resources examined are televisions, radios, satellite dishes (for television), and cars. Nearly every household has at least one television and radio in all governorates. Satellite dishes are relatively rare but don't seem to follow the same trends that we've seen for other resources with correlations to consumption and whether or not the governorate is coastal. Du kef, for example, has a decently high proportion of satellite dishes when compared to other governorates, despite having low total levels of leisure and ICT resources. Results for differences by car are a little more difficult to interpret, and I believe there are many factors at play there including how urban and driveable a governorate is, availability of public transportation, and whether people need to drive cars for work.

Finally, we take a look at how luminosity has changed over time in order to gleam insights about how they may reveal changes in Tunisia economic activity throughout the years. Note that we previously could not look at how luminosity compared to governorate-level economic indicators such as household appliances, leisure resources, and consumption over time because of data limitations, namely that those metrics have not been been consistently measured throughout the years in Tunisia.

### (4.g) Luminosity & the Tunisian Economy, 1992-2013

```r
lum <- read_csv("../data/final/tun_lum_governorate_93_13_tidy.csv")

# Plot of Luminosity by Governorate by Year
ggplot(lum, aes(x = year, y = mean, colour=NAME_1)) +
geom_line() +

xlab("Year") +
ylab("Luminosity (watts/cm\u00b2)") +
labs(color='Governorate',
    caption = "Source: L'Institut National de la Statistique (INS)") +
    ggtitle("Luminosity by Governorate by Year",
        subtitle = "1992 - 2013") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```

Luminosity by Governorate by Year
1992 – 2013
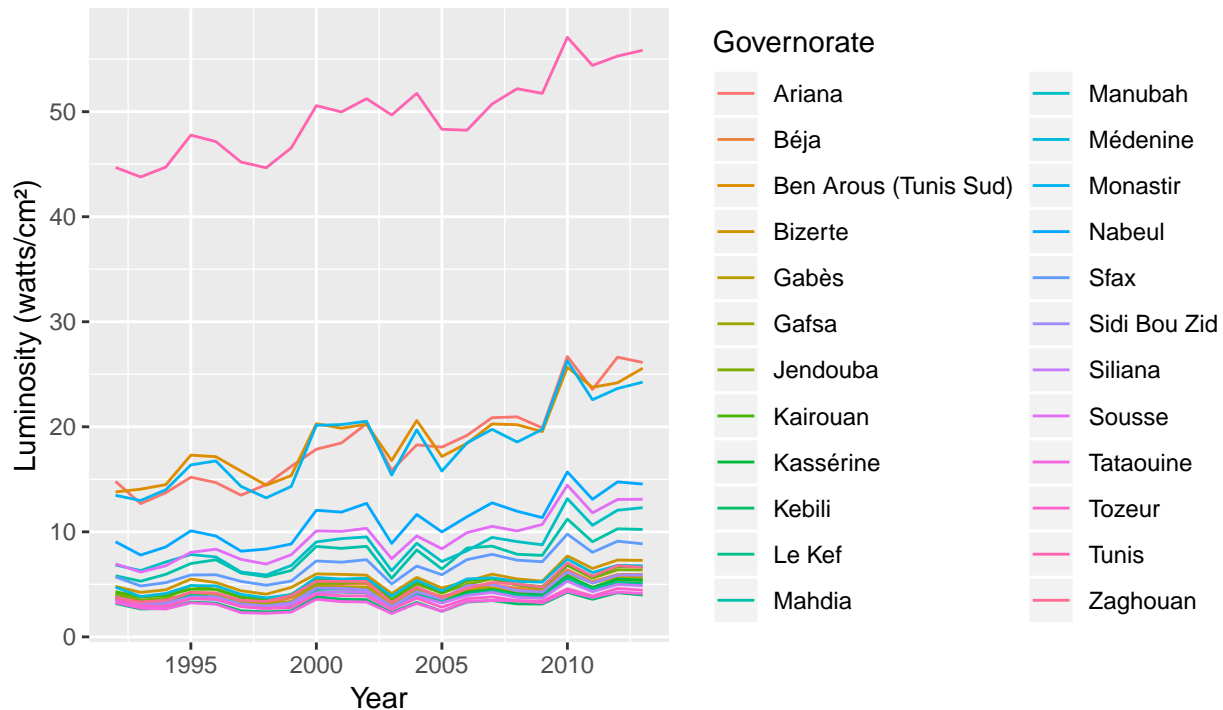
Source: L'Institut National de la Statistique (INS)

This plot of luminosity by governorate by year shows that Tunis has had a significantly higher level of luminosity than the other governorates. The three governorates in the middle of the graph (Ariana, Ben Arous, and Monastir) are suburban extensions of Tunis. The rest of the governorates are bunched together at the bottom of the plot. In general there is an upward trend in luminosity from 1992 to 2013, but it is not consistent from year to year, as there are peaks and valleys that show that the increase in light is not linear. For an additional frame of reference, see the appendix for an illustrative example of a few governorates with high, medium, and low levels of luminosity from 1992 to 2013.

Of particular interest is the sharp downward spike in luminosity for every governorate from 2010 to 2011. 2011 is the year of the Tunisian Revolution, and it is very possible that the political and economic turmoil of that revolution caused luminosity in Tunisia to crash for that particular year.

```
tun <- lum %>%
   group_by(year) %>%
   summarise(count = sum(count), sum = sum(sum))

tunisia_economy <- read_csv("../data/intermediate/tunisia_economy.csv")
tunisia <- merge(tun,tunisia_economy,by="year")

tunisia$avg_lum <- tunisia$sum / tunisia$count

melt_tunisia <- melt(tunisia,id="year")

melt_tunisia <- melt_tunisia %>%
     group_by(variable) %>%
     arrange(year) %>%
     mutate(indexedvalue = 100 * value / first(value))
```

```
ggplot(subset(melt_tunisia, variable %in% c("avg_lum", "govt_debt", "gdp_per_capita", "gdp_growth", "in
      aes(x=year,
      y=indexedvalue,
      color=variable, group = variable))+
  geom_line() + ggtitle("Indexed Economic Indicators for Tunisia", subtitle = "1992 - 2013") +
  scale_color_discrete(name = "Variable", labels = c("GDP Growth Rate", "GDP per capita (PPP)", "Inflat
  xlab("Year") +
  ylab("Indexed Value") +
  labs(color='Variable',
caption = "Sources: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```

Indexed Economic Indicators for Tunisia

1992 – 2013
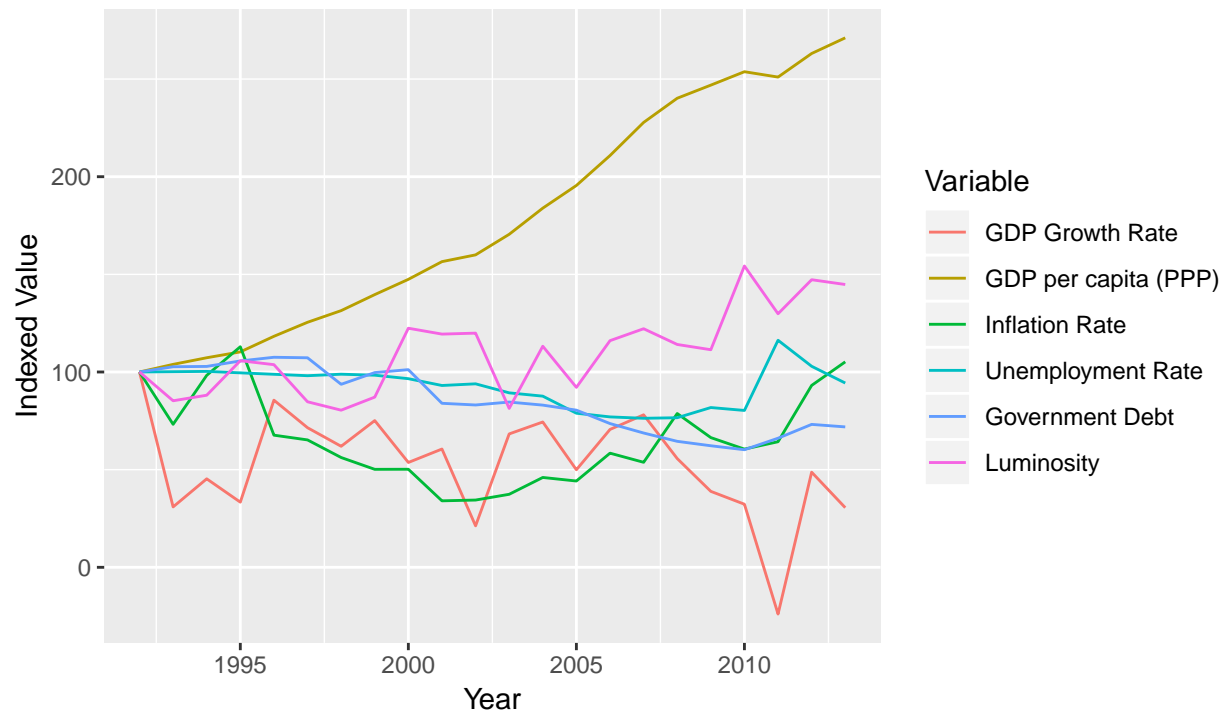
s: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database

This plot further explores the connection between luminosity and Tunisian economic indicators from 1992 to 2013, but this time on a national level. Note that each variable is indexed to the value from the base year of 1992 and at the starting value of 100.

'Luminosity' in this plot is the average level of luminosity in watts/cm2 for the entire country of Tunisia from 1992 to 2013. This is calculated by summing up the total sum of pixel values of luminosity for all of the governorates as well the total count of the number of pixels for each governorate (which is essentially a static number of the amount of pixels that comprise a given land area) for all of Tunisia, then dividing the total sum by the total count for each year from 1992 to 2013.

As we have noted before, while there is a general upward trend in luminosity from 1992 to 2013, it is not consistent from year to year. In contrast GDP per capita has increased every single year since 1992 with the exception of 2011. In 2011, both GDP growth and GDP per capita sharply fall, while luminosity, inflation,

the unemployment rate, and government debt all increase to varying degrees. Overall 2011 was a very bad year for Tunisia economically.

However in this graph it is difficult to ascertain a particular relationship between luminosity and other economic indicators. Because of this we decided to perform a correlation test to see how correlated the level of luminosity in Tunisia was to the other variables.

```
cor_tunisia <- cor(tunisia)
```

```
## Warning in cor(tunisia): the standard deviation is zero
```

```
cor_tunisia
```

```
##                        year count        sum gdp_growth       gdp_ppp
## year              1.0000000    NA  0.7329709 -0.4007986  0.989903076
## count                    NA     1         NA         NA           NA
## sum               0.7329709    NA  1.0000000 -0.3818728  0.751756723
## gdp_growth       -0.4007986    NA -0.3818728  1.0000000 -0.397664279
## gdp_ppp           0.9899031    NA  0.7517567 -0.3976643  1.000000000
## gdp_per_capita    0.9895421    NA  0.7433196 -0.3841289  0.999346158
## inflation        -0.1098110    NA  0.1436795 -0.1029147 -0.008166219
## unemployment_rate -0.3512514    NA -0.1185138 -0.3242774 -0.355587641
## govt_debt        -0.9141032    NA -0.6580060  0.3674089 -0.927997907
## avg_lum           0.7329709    NA  1.0000000 -0.3818728  0.751756723
##                   gdp_per_capita    inflation unemployment_rate   govt_debt
## year                  0.98954214 -0.109810976        -0.3512514  -0.9141032
## count                         NA           NA                NA          NA
## sum                   0.74331962  0.143679478        -0.1185138  -0.6580060
## gdp_growth           -0.38412887 -0.102914738        -0.3242774   0.3674089
## gdp_ppp               0.99934616 -0.008166219        -0.3555876  -0.9279979
## gdp_per_capita        1.00000000 -0.027981867        -0.3835113  -0.9366557
## inflation            -0.02798187  1.000000000         0.3097474   0.1539523
## unemployment_rate    -0.38351131  0.309747374         1.0000000   0.5039443
## govt_debt            -0.93665573  0.153952295         0.5039443   1.0000000
## avg_lum               0.74331962  0.143679478        -0.1185138  -0.6580060
##                      avg_lum
## year               0.7329709
## count                     NA
## sum                1.0000000
## gdp_growth        -0.3818728
## gdp_ppp            0.7517567
## gdp_per_capita     0.7433196
## inflation          0.1436795
## unemployment_rate -0.1185138
## govt_debt         -0.6580060
## avg_lum            1.0000000
```

The correlation of luminosity with total GDP is 0.75 and with GDP per capita is 0.74, which suggests that luminosity can be a fairly strong proxy of measures of GDP. While there is little correlation with measures of inflation (0.14) and unemployment rate (-0.11), they are at least in the direction that we expect – both higher levels of inflation and luminosity would suggest greater levels of economic activity, while unemployment rates would likely have a negative relationship with luminosity, as high levels of unemployment suggest the economy is not doing well and that there is less light to be observed in a given year.

Of particular interest is that the relationship between luminosity and government debt is very strong in the negative direction (-0.66). A possible explanation for this is that higher levels of government debt are associated with the economy not doing well, and thus there is less luminosity to be observed.

```r
ggplot() +
geom_line(data=tunisia,aes(y = rescale(avg_lum),x= year,colour="darkblue"),size=1.5, alpha =.5)+
  geom_line(data=tunisia,aes(y = rescale(gdp_ppp), x=year,colour="red"),size=1.5, alpha =.5)+
  scale_color_discrete(name = "Normalized Variable", labels = c("Luminosity", "GDP per capita (PPP)")) +
    ggtitle("Luminosity vs. GDP per capita (PPP) in Tunisia (Normalized)",
        subtitle = "1992 - 2013") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68")) + xlab("Year") +
  ylab("Normalized Value") + labs(color='Normalized Value',
 caption = "Sources: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database")
```

Luminosity vs. GDP per capita (PPP) in Tunisia (Normalized)
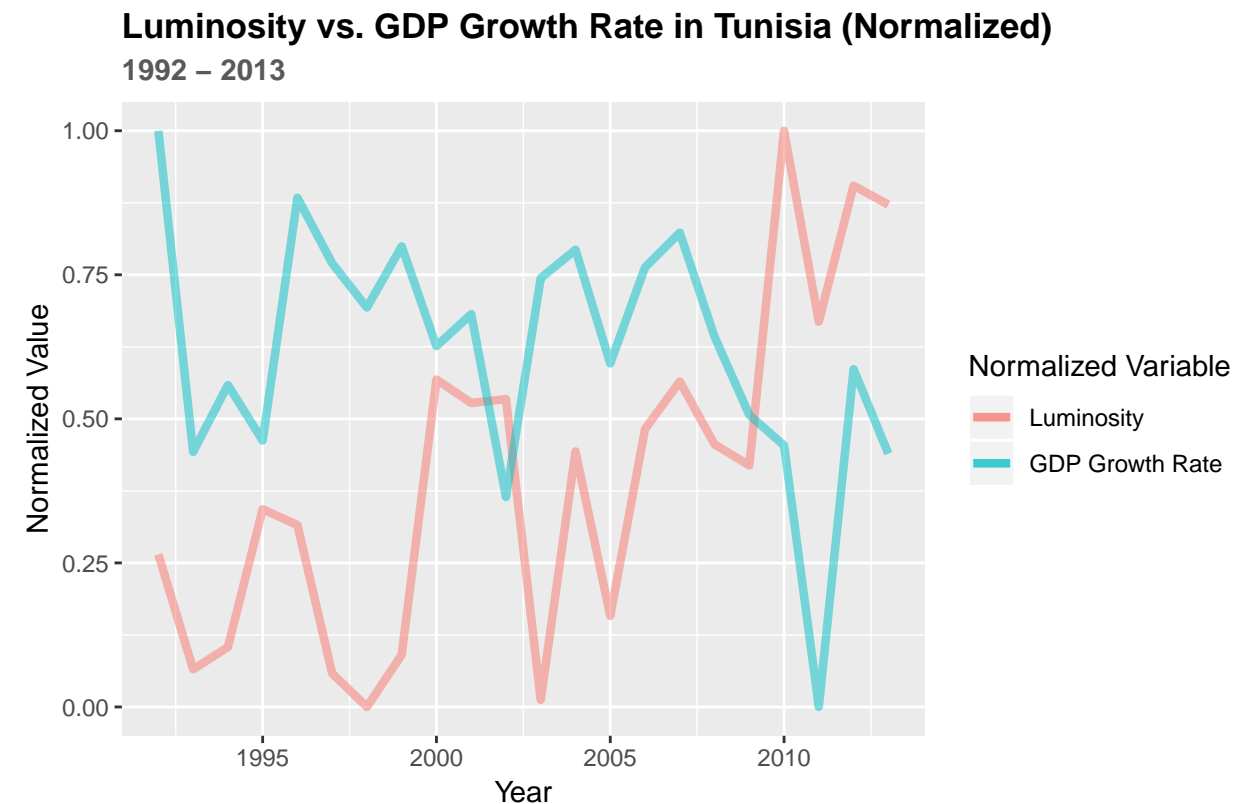**1992 – 2013**
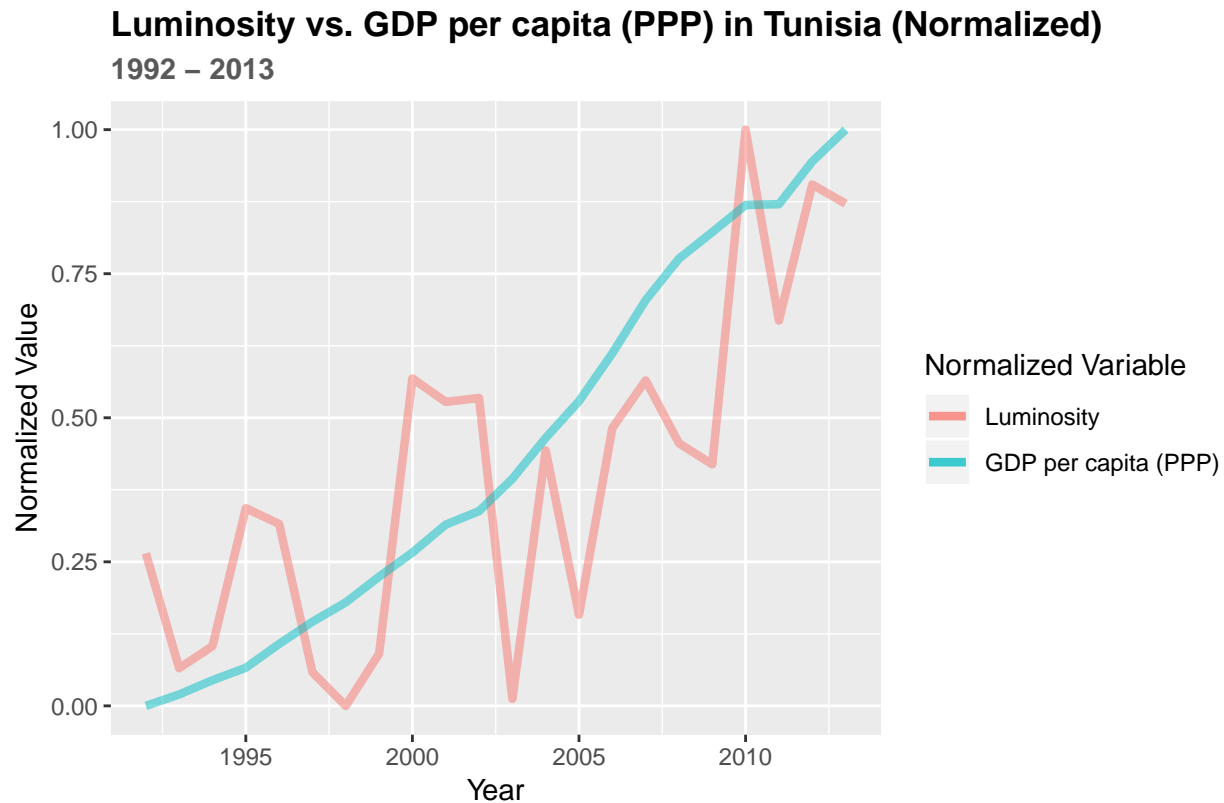
s: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database

This graph shows the normalized relationship between the level of luminosity and GDP per capita PPP in Tunisia from 1992 to 2013. While the correlation between the two measures is 0.74 and both show an overall upward trend during the given time interval, we can see that GDP per capita (PPP) is a relatively stable measure, while the level of luminosity wildly fluctuates from year to year. This suggests that GDP per capita is a more reliable measure of economic activity compared to the level of luminosity.

```r
ggplot() +
geom_line(data=tunisia,aes(y = rescale(avg_lum),x= year,colour="darkblue"),size=1.5, alpha =.5)+
  geom_line(data=tunisia,aes(y = rescale(gdp_growth), x=year,colour="orange"),size=1.5, alpha =.5)+
  scale_color_discrete(name = "Normalized Variable", labels = c("Luminosity", "GDP Growth Rate")) +
    ggtitle("Luminosity vs. GDP Growth Rate in Tunisia (Normalized)",
        subtitle = "1992 - 2013") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68")) + xlab("Year") +
```
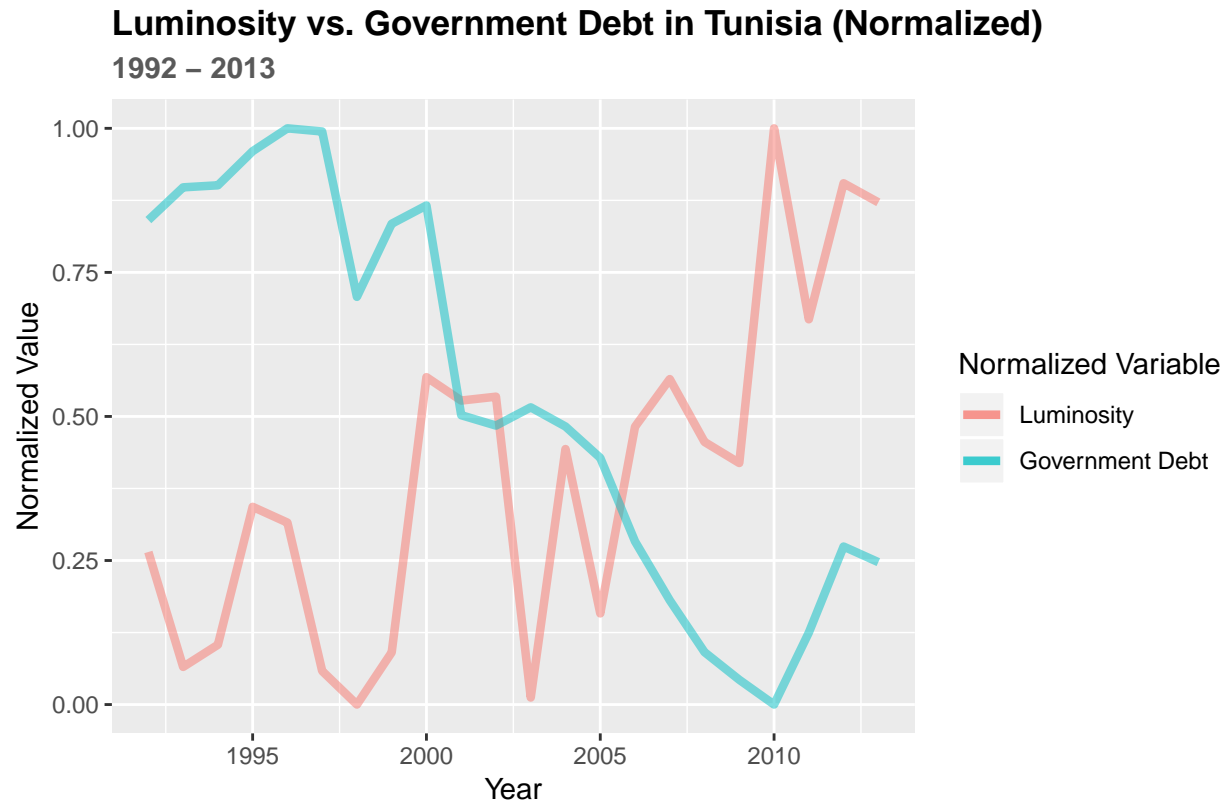
```
    ylab("Normalized Value") + labs(color='Normalized Value',
caption = "Sources: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database")
```

Luminosity vs. GDP Growth Rate in Tunisia (Normalized)

1992 – 2013

Recall that luminosity and GDP growth have a slightly negative relationship (-0.38). This may provide some evidence for the theory of economic convergence, namely that poorer regions (that we assume to be lower in luminosity) grow at faster rates than more developed regions. This graph of the relationship between the two variables shows that when normalized luminosity is low, normalized gdp growth tends to be high and vice versa; observe the time ranges of 1995-2000, 2003 to 2008, and 2011-2013 to see this trend.

```
ggplot() +
geom_line(data=tunisia,aes(y = rescale(avg_lum),x= year,colour="darkblue"),size=1.5, alpha =.5)+
  geom_line(data=tunisia,aes(y = rescale(govt_debt), x=year,colour="red"),size=1.5, alpha =.5)+
  scale_color_discrete(name = "Normalized Variable", labels = c("Luminosity", "Government Debt")) +
  ggtitle("Luminosity vs. Government Debt in Tunisia (Normalized)", subtitle = "1992 - 2013") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68")) + xlab("Year") +
  ylab("Normalized Value") + labs(color='Normalized Value',
  caption = "Sources: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database")
```

Luminosity vs. Government Debt in Tunisia (Normalized)
1992 – 2013

Finally, the negative relationship between government debt and luminosity can be seen in this graph. From 1992 to 2000, normalized government debt is high while normalized luminosity is low. From 2000 onwards, government debt begins to steadily decrease until 2011 when it sharply spikes again. Luminosity generally seems to climb over this timespan, but again in a very non-linear way. Overall, it is clear from these graphs that even as luminosity generally increases over time, it fluctuates wildly from year to year, much more so than other, more official economic indicators that are measured across time.

(5) Executive summary

We began this project with the intention of exposing a connection between luminosity in nighttime satellite images and economic attributes within the country of Tunisia. We decided to look at the country both as a whole and by governorate, an administrative boundary which serves a similar function as a state in the United States. We began by examining how these governorates differ from one another. The first thing we chose to examine was population density. We had a hypothesis that more densely populated governorates would have higher luminosity, so we wanted to identify which governorates to focus on for these analyses.

```r
admin1_choropleth("tunisia", density, num_colors = 7, title = "Population Density", legend = "Population
  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"))
```

Tunisia's Population Density by Governorate
Year: 2014

Population Density (km²)

- [3.84 to 48.10)
- [48.06 to 52.30)
- [52.31 to 63.90)
- [63.93 to 129.40)
- [129.43 to 252.90)
- [252.89 to 538.60)
- [538.59 to 3,052.70]
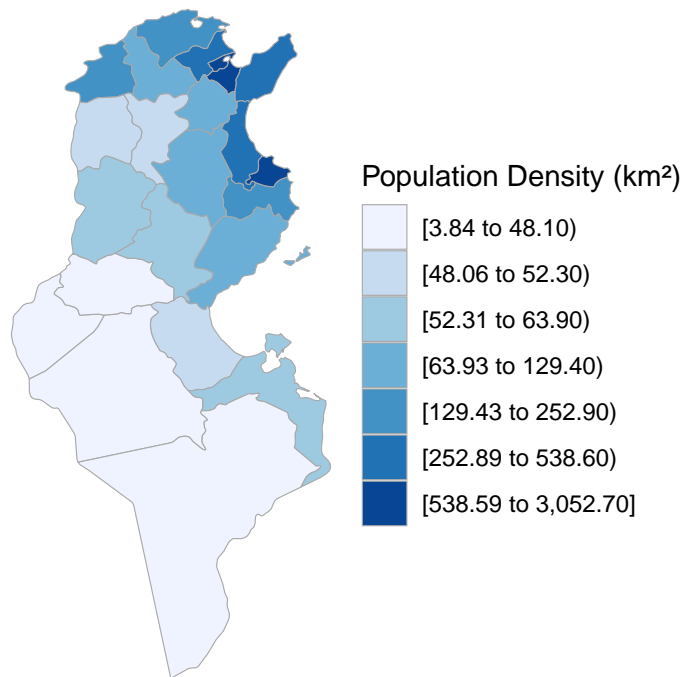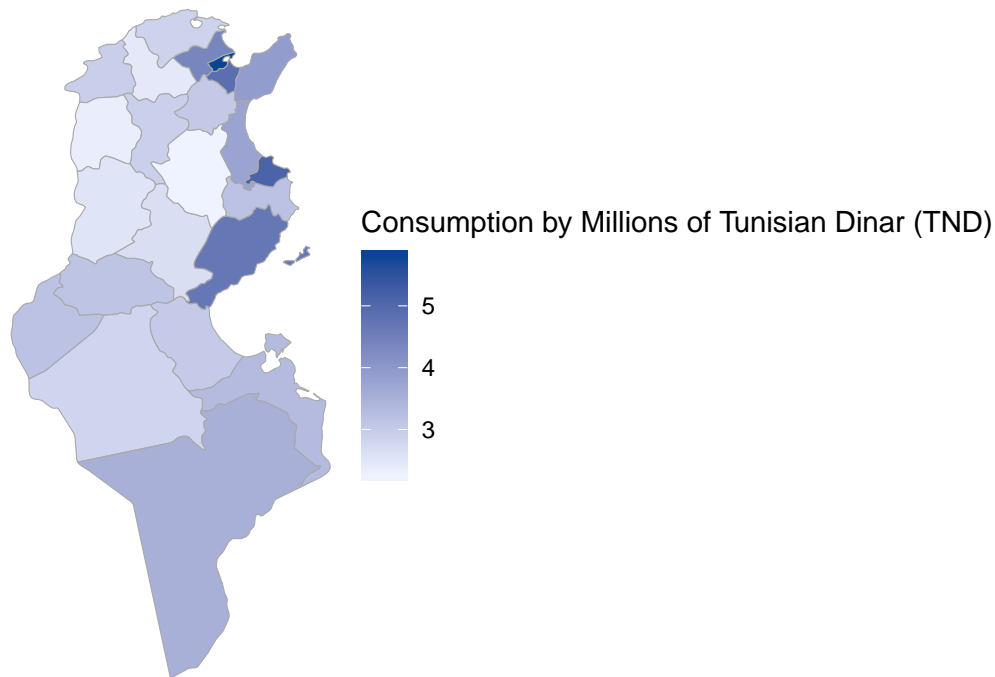
Source: L'Institut National de la Statistique (INS)

This graph gave us our first indications that if luminosity and population density are inter-related like we hypothesized that we may want to pay special attention to governorates in coastal regions, which tend to have a higher population density.

The other variable we expected to be strongly correlated with luminosity is consumption per capita. Consumption measures how much money people are spending, and is an indicator of a strong economy.

```
admin1_region_choropleth(df, num_colors= 1, legend = "Consumption by Millions of 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"))
```

Consumption by Governorate

Year: 2015

Consumption by Millions of Tunisian Dinar (TND)

5

4

3

This graph showed that while the same 3 governorates (Tunis, Monastir and Ben Arous) have the highest values of both consumption per capita and population density, there are some differences. A few larger coastal governorates have low consumption levels and a few larger, low population density landlocked governorates have high rates of consumption per capita.

After learning these lessons about Tunisia, we shifted our focus to the luminosity data:

```
# Plot of Luminosity by Governorate by Years
#subset(lum, NAME_1 %in% c("Tunis", "Sfax", "Monastir", "Tataouine")
ggplot(subset(lum, NAME_1 %in% c("Tunis", "Sfax", "Monastir", "Tataouine", "Le Kef", "Sidi Bou Zid", "Mc
                                  "Ben Arous (Tunis Sud)")),
       aes(x = year, y = mean, colour=NAME_1)) +
geom_line() +
scale_colour_brewer(palette="Set3") +
xlab("Year") +
ylab("Luminosity (watts/cm\u00b2)") +
labs(color='Governorate', caption = "Source: L'Institut National de la Statistique (INS)") +
   ggtitle("Luminosity by Governorate by Year",
           subtitle = "1992 - 2013") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68"))
```

Luminosity by Governorate by Year
1992 – 2013
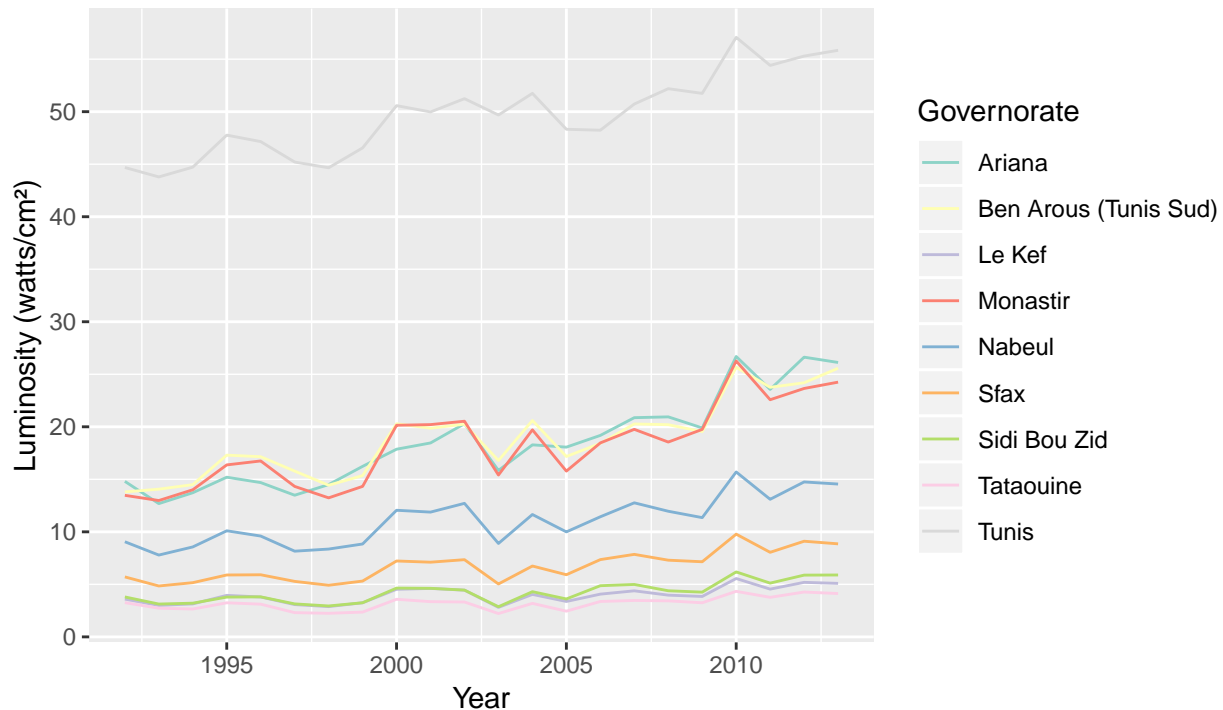
Source: L'Institut National de la Statistique (INS)

This shows luminosity data over time for several select governorates. While a number of factors make it difficult for us to make year-to-year comparisons for satellite luminosity data, from this chart we can observe the following:

- Tunis, as would be expected from its high levels of population density and consumption per capita, has the highest levels of luminosity by a wide margin. Population density and consumption per capita also seem to generally be relatively good predictors of luminosity.
- Luminosity collected in this manner and viewed as a time series is noisy. In general it trends upwards over time, but the relationship isn't smooth and simple.
- Although luminosity is unstable from year to year, the governorates generally seem to be subject to these sharp yearly changes in very similar ways. There does not appear to be a lot of variation in luminosity between the governorates from year to year.
- There is a very sharp downward spike in luminosity in 2011 that coincides with the Tunisian Revolution. It is possible that the political and economic turmoil of that revolution had a big negative shock on Tunisia's luminosity for that year.

Our final aim was to draw a connection between various economic indicators and luminosity in Tunisia over time:

```
ggplot() +
geom_line(data=tunisia,aes(y = rescale(avg_lum),x= year,colour="darkblue"),size=1.5, alpha =.5)+
  geom_line(data=tunisia,aes(y = rescale(govt_debt), x=year,colour="red"),size=1.5, alpha =.5)+
  scale_color_discrete(name = "Normalized Variable", labels = c("Luminosity", "Government Debt")) +
  ggtitle("Luminosity vs. Government Debt in Tunisia (Normalized)", subtitle = "1992 - 2013") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "bold", color = "grey35")) +
  theme(plot.caption = element_text(color = "grey68")) + xlab("Year") +ylab("Normalized Value") + labs(
```

Luminosity vs. Government Debt in Tunisia (Normalized)
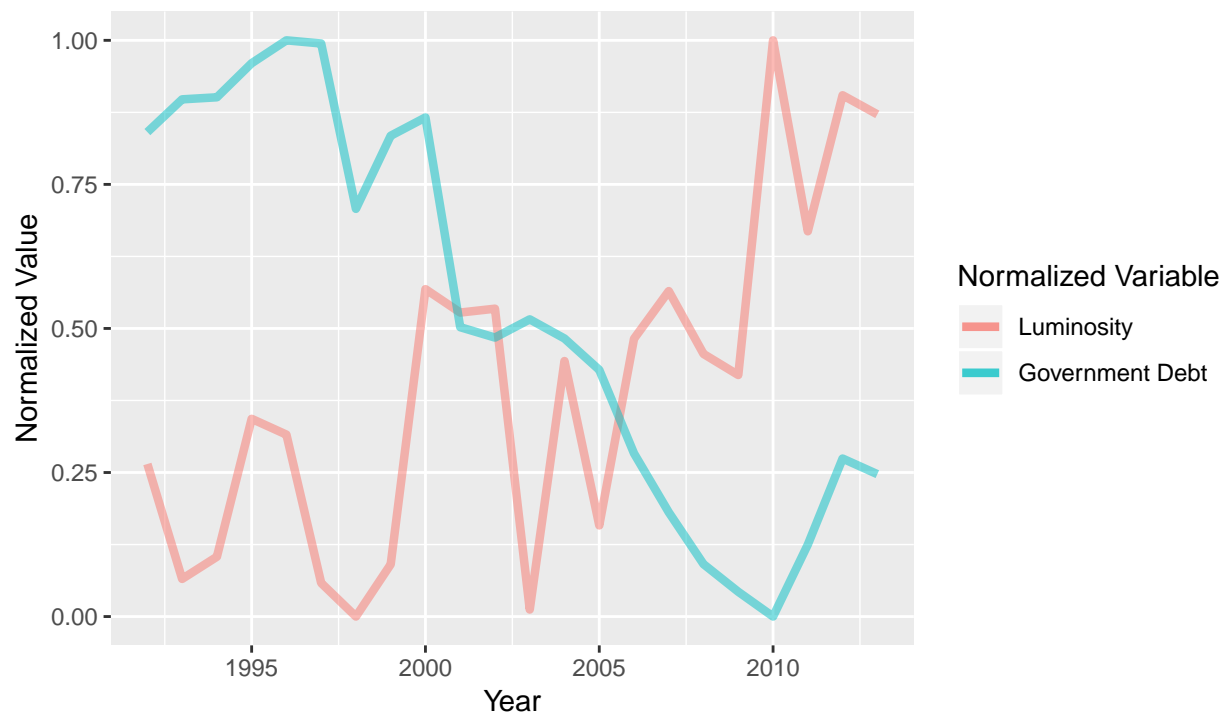
1992 – 2013

This graph shows one of our most interesting findings. It compares luminosity (watts/ cm2) and government debt (TND), with both values scaled from 0 to 1 to show effects despite the two being on different scales. It illustrates (1) just how unstable luminosity is from year to year, even as it has generally gone up in the long run and (2) a fairly strong inverse relationship between luminosity and government debt. From 1992 to 2000, normalized government debt is high while normalized luminosity is low. From 2000 onwards, government debt begins to steadily decrease until 2011 (again, the year of the Tunisian Revolution) when it sharply spikes again. Luminosity increases from 2001 to 2013, but again in a very non-linear way. Overall, the main takeaway from this graph is that even as luminosity generally increases over time, it fluctuates wildly from year to year, much more so than other panel data economic indicators such as government debt.

(6) Interactive component

Since our report provides detailed plots pertaining to economic factors, we wanted our interactive component of the project to focus on visualizing data that is hard to map at once; that is luminosity over time. For instance, trying to facet graphs covering 22 years worth of data is visually cumbersome. Instead, we are giving our user the liberty to explore luminosity using a slider. We specifically chose the variable, luminosity, as it is the central and more importantly the novel piece in our report. Producing luminosity data required us to process NASA satellite imagery as the information is, as of now, not readily available.

Shiny App URL:

https://nightlight.shinyapps.io/TunisiaNightLightOverTime/

You may also visualize an animation of the original nightlight satellite imagery via the below link. Because it is hard to discern the subtle differences from year to year from the satellite images, we decided to instead create a fast looping GIF to better represent contrast. https://github.com/mariemayadi/data/blob/master/luminosityOverTimeTunisia.gif

Note: The app is experiencing some latency issues. We apologize for the inconvenience.

(7) Conclusion

We chose to perform this analysis on Tunisia rather than the United States or any other country because as a group we were inspired by our teammate's personal connection to the country. Although historically there has been some work done to connect luminosity data to economic indicators, developing countries like Tunisia and its neighbors have often been neglected in these types of analyses.

As one would be expect, mean luminosity over the country in general increased over time from 1992 to 2013. However, due to the satellite data limitations discussed above, the trend is subject to sudden changes that may or may not reflect actual luminosity changes. This makes it difficult to look at luminosity longitudinally.

Perhaps the largest connections between luminosity and other data were made based on consumption per capita (a measure of economic activity) and geographic location. Coastal governorates tend to have higher consumption and luminosity. This is likely because these governorates contain established port cities such as Tunis and Sfax, where trade brings in money and increases consumption. We believe this is the primary driving force behind luminosity.

We also compared changes in luminosity over time to economic indicators such as GDP per capita, GDP growth, and government debt on a national level from 1992 to 2013. We found that while luminosity has fairly strong relationships with some of these economic indicators, as a variable it is extremely unstable from year to year. This suggests that it would not be that reliable of a measure in any given year, but when aggregated over longer timespans can be as representative of Tunisia's economic performance as other measures.

With that being said, there was is a lot of data that we didn't have time to analyze in as much depth as we'd have preferred, and some we couldn't get to at all. We'd also have preferred to have some additional economic indicators per governorate, such as median household income or GDP per capita. We'd also like to do more direct comparisons between some of the data from the Tunisia Data Portal (resources, employment industry) and luminosity. And despite having many variables to explore, most of them only had data for 2014, which was after our luminosity data ended in 2013. It would be preferable to have this data in time series and by governorate as well, so we could look at luminosity along with these variables together longitudinally and on a subnational level.

One of our most important lessons learned during the course of this project is to be skeptical of our biases and preconceived notions when analyzing a dataset, as we were often surprised by what we found. For instance, the high rates of women working manufacturing jobs in some governorates, and the relatively high consumption of some larger more desertous governorates such as Tataouine, which was the inspiration for Luke Skywalker's lawless backwater planet in Star Wars. We also learned how US-centric many online tools and resources can be. It is much easier to create a map of the United States than any other country using R. That being said, there are people working hard to extend resources to cover less commonly examined regions. ChoroplethrAdmin1, despite its limitations (it is missing a governorate and doesn't support French language well) is an incredibly useful tool. The Database of Global Administrative areas from which we sourced the shapefiles for Tunisian governorates and delegations also comes to mind. We believe that in shining a light on Tunisia in this manner we have learned several important lessons and made good use of data that historically has not gotten the attention that it deserves.

Citations

Chen, Xi, and William D. Nordhaus. The value of luminosity data as a proxy for economic statistics. No. w16317. National Bureau of Economic Research, 2010.

Elvidge, Christopher D., et al. "Potential for global mapping of development via a nightsat mission." GeoJournal 69.1-2 (2007): 45-53.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. "Measuring economic growth from outer space." American economic review 102.2 (2012): 994-1028.

Michalopoulos, Stelios, and Elias Papaioannou. "Pre-colonial ethnic institutions and contemporary African development." Econometrica 81.1 (2013): 113-152.

Appendix

Section I

While we believe the stacked bar charts to be a better representation to compare total numbers of resources by governorate, it can sometimes be difficult to interpret data values in isolation. For this reason we also included a faceted Cleveland dot plot for easier direct lookup of data for an individual governorate.

```
g_dot_facet <- ggplot(appliances_tidy, aes(x = appliances_household,
                                    y = reorder(Region, appliances_household))) +
    geom_point() + ylab("") +
    ggtitle("appliances Resources per Household")


g_dot_facet + facet_wrap(~appliances_type, ncol=1)
```
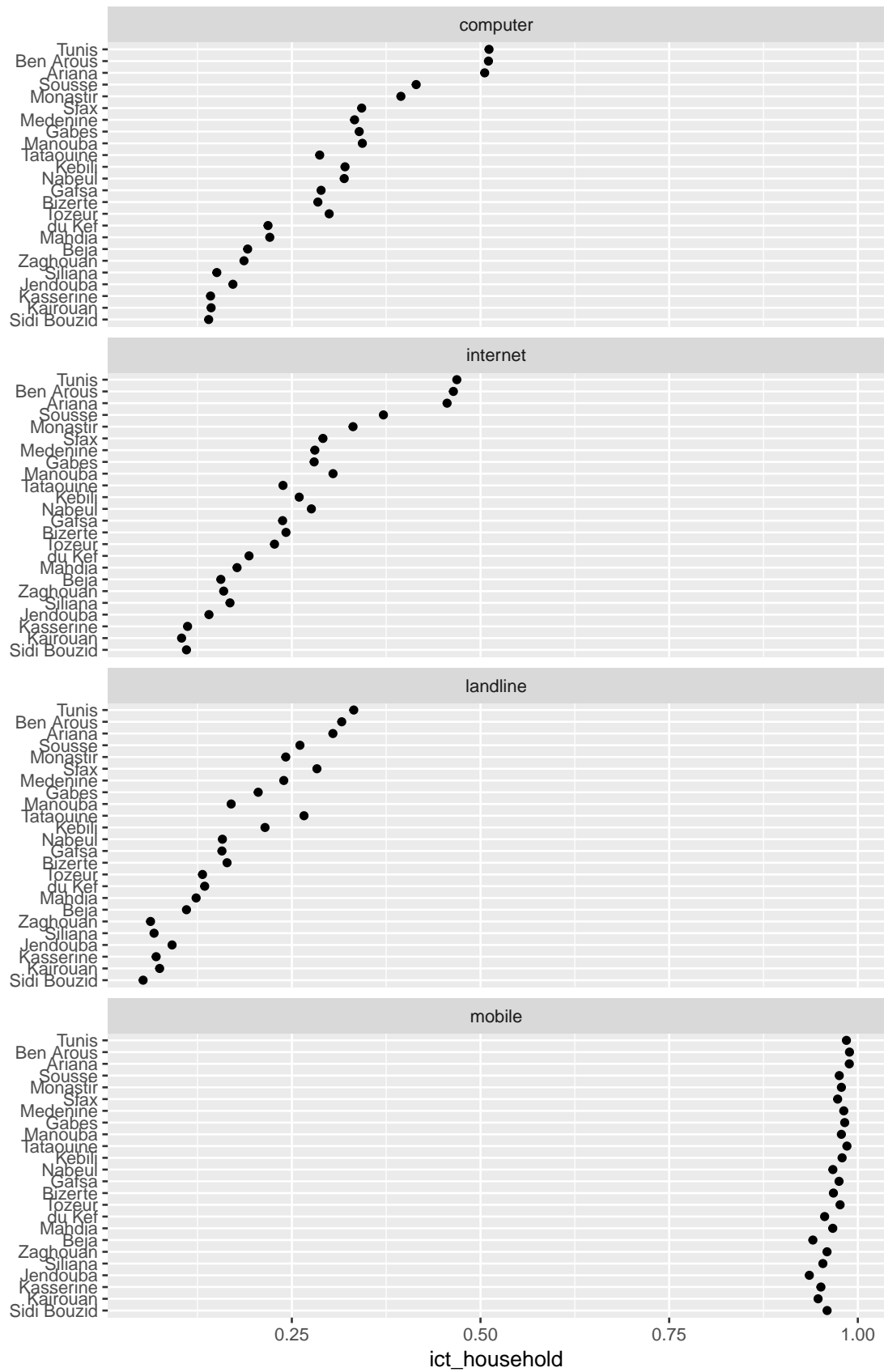
appliances Resources per Household
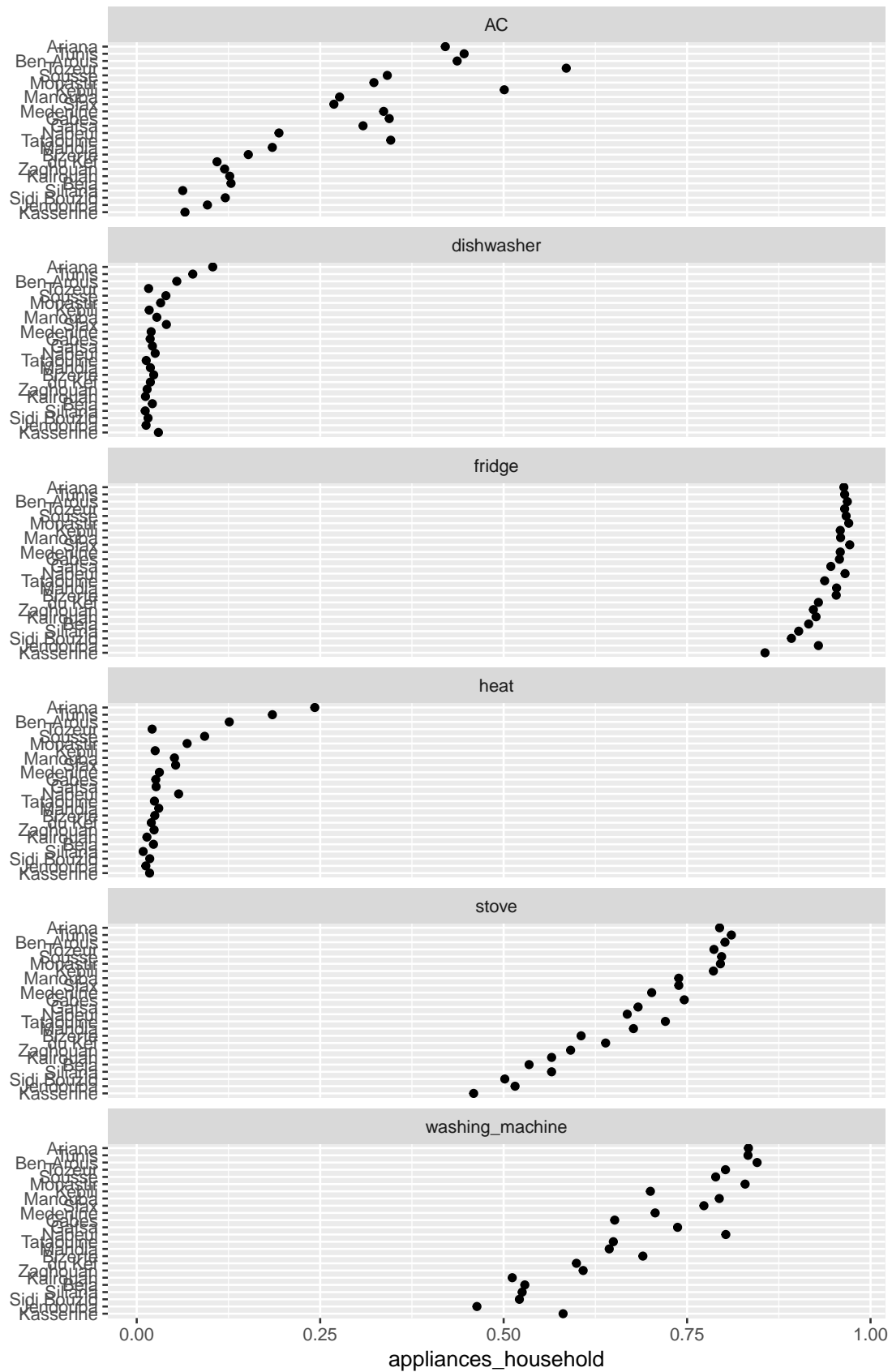
```r
g_dot_facet <- ggplot(internet_tidy, aes(x = ict_household,
                                 y = reorder(Region, ict_household))) +
    geom_point() + ylab("") +
    ggtitle("internet Resources per Household")

g_dot_facet + facet_wrap(~ict_type, ncol=1)
```

internet Resources per Household

```
g_dot_facet <- ggplot(leisure_tidy, aes(x = leisure_household,
                                         y = reorder(Region, leisure_household))) +
    geom_point() + ylab("") +
    ggtitle("Leisure Resources per Household")

g_dot_facet + facet_wrap(~leisure_type, ncol=1)
```

Leisure Resources per Household
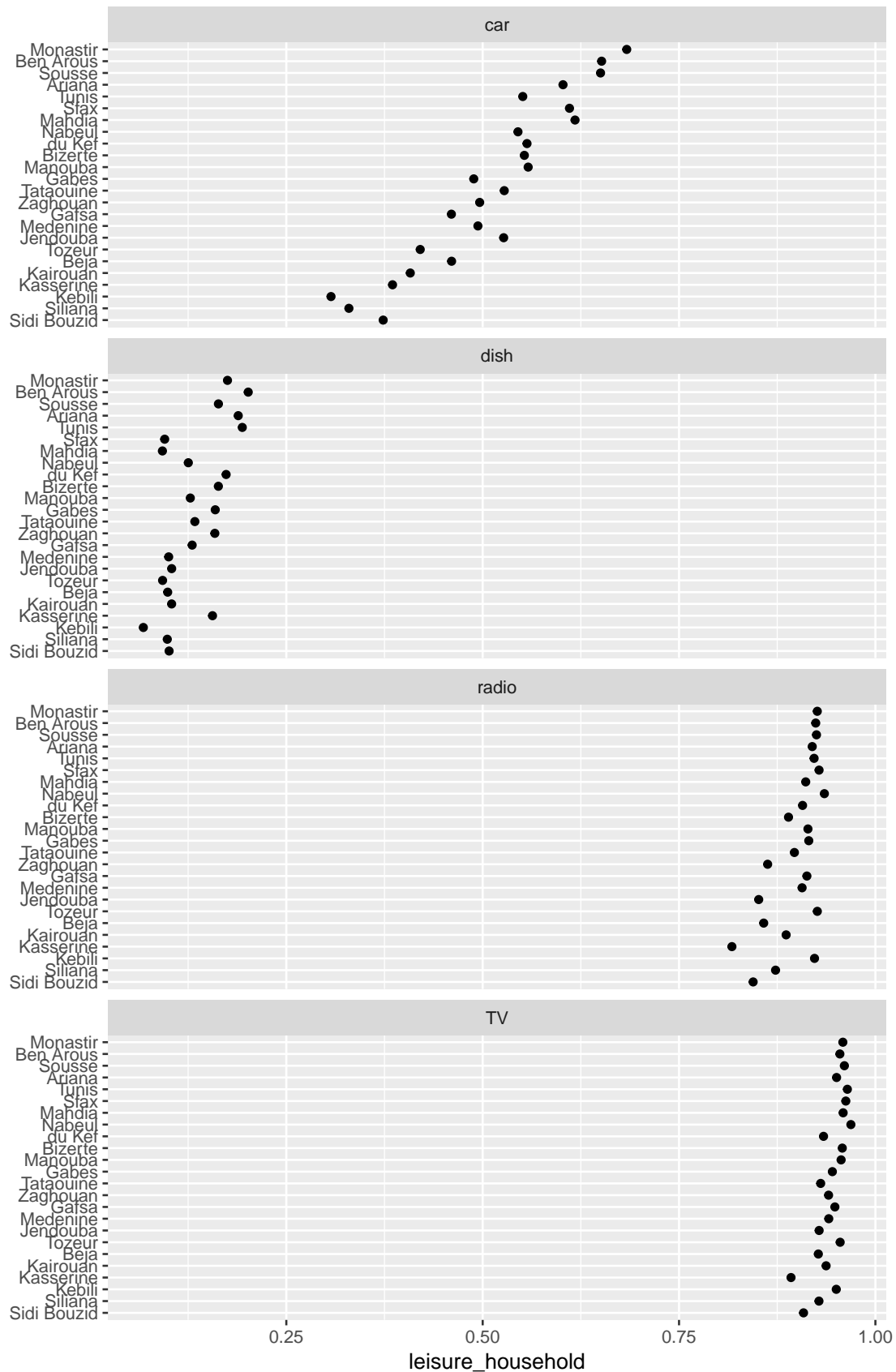
```r
# Subset of Data
ggplot(subset(lum, NAME_1 %in% c("Tunis", "Sfax", "Monastir", "Tataouine")),
       aes(x=year,
       y=mean,
       color=NAME_1, group = NAME_1))+
  geom_line() +
ylab("Luminosity") +
ggtitle("Luminosity of Four Governorates by Year") + labs(color='Governorate')
```

Luminosity of Four Governorates by Year
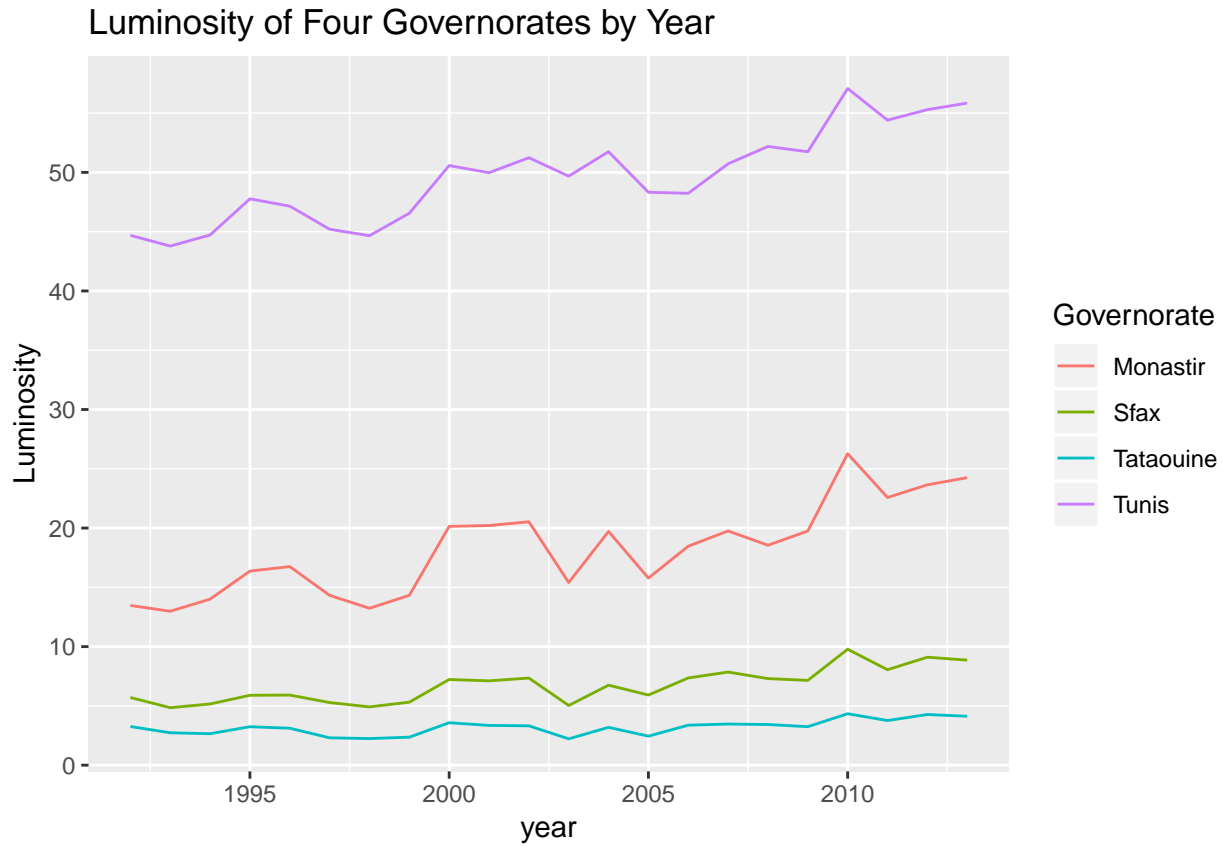
This is a representative subset of high, medium, and low level luminosity governorates as well as Sfax. From this we can see that the governorates are at completely different levels, but actually move fairly similarly throughout the years; overall if one governorate increases in luminosity for a given year, from this graph it is very reasonable to infer that the other governorates will increase in luminosity as well (and vice versa for decreases in luminosity). The Sfax and Tatouine governorates in particular almost seem to operate in unison, with Sfax having a slightly higher level of luminosity throughout but with their time-series lines over time looking very similar.
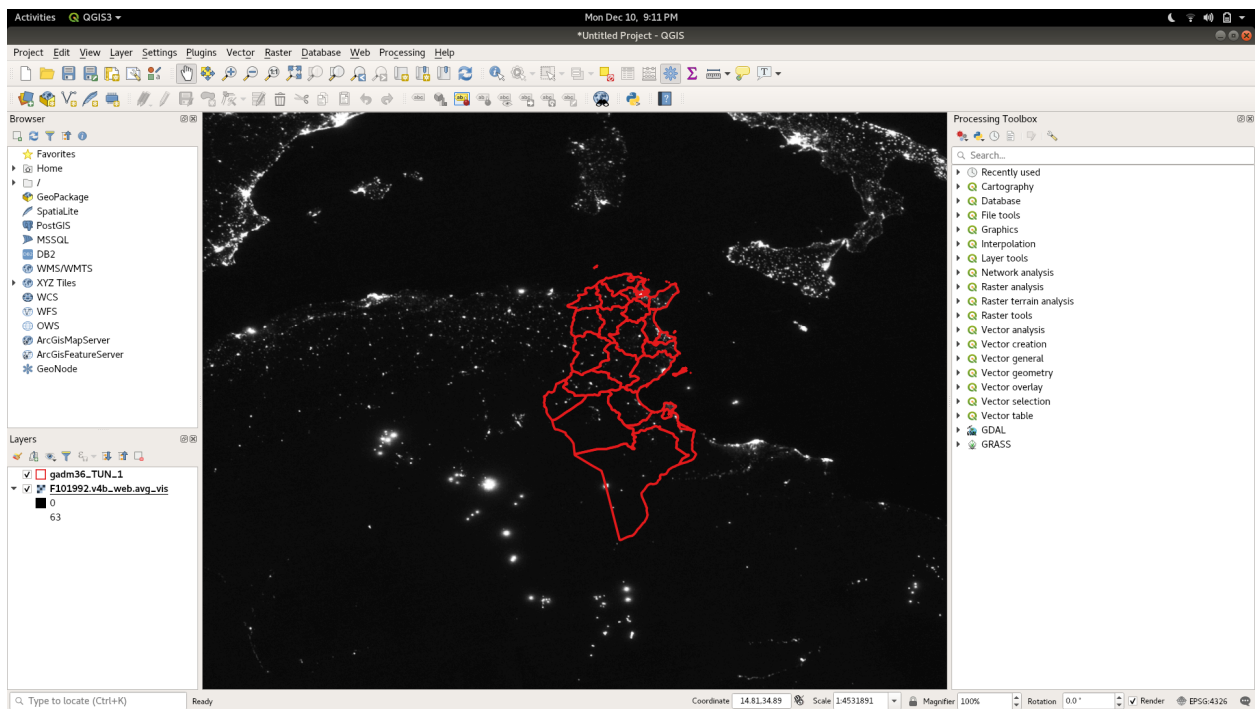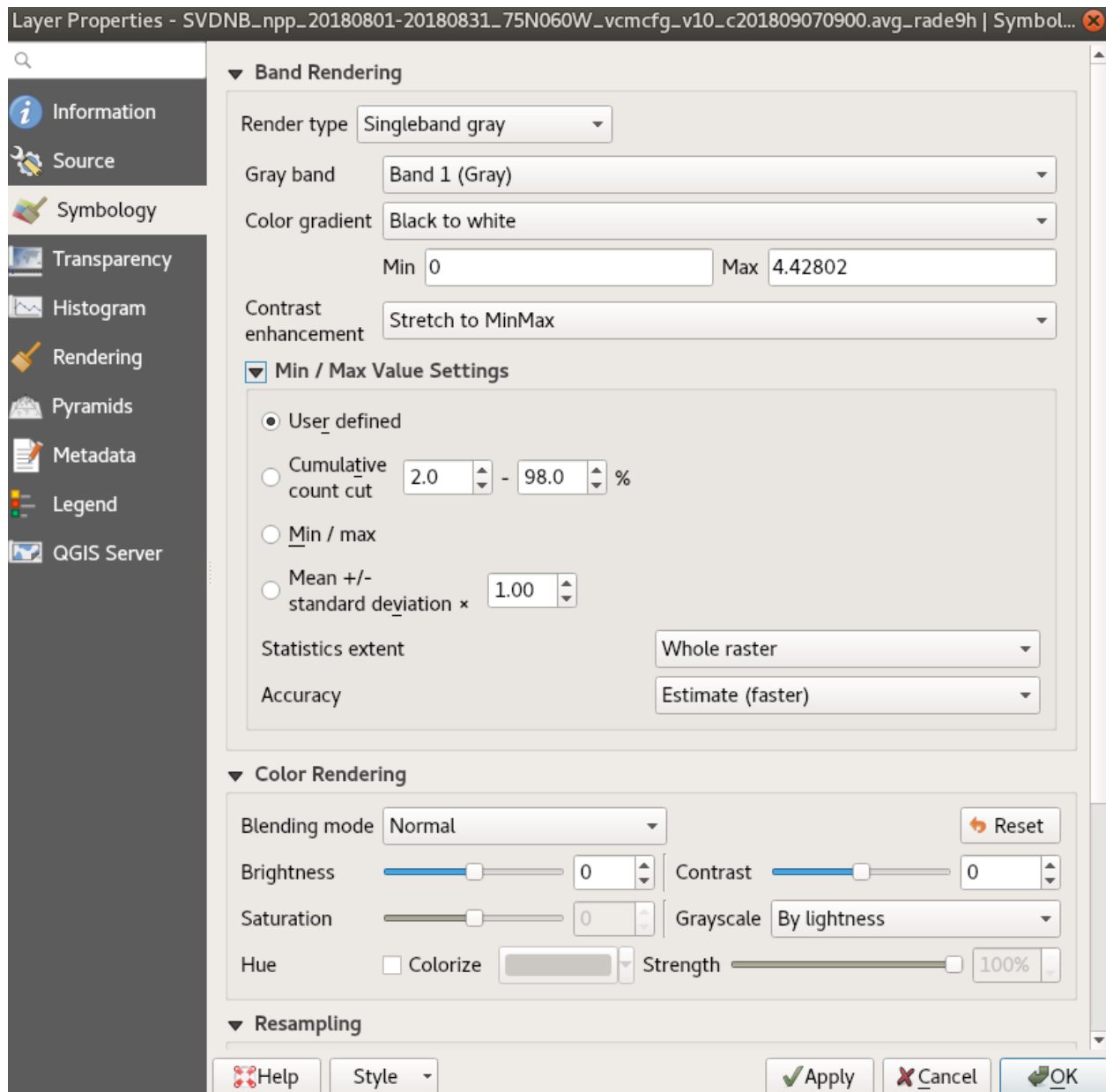
Section II

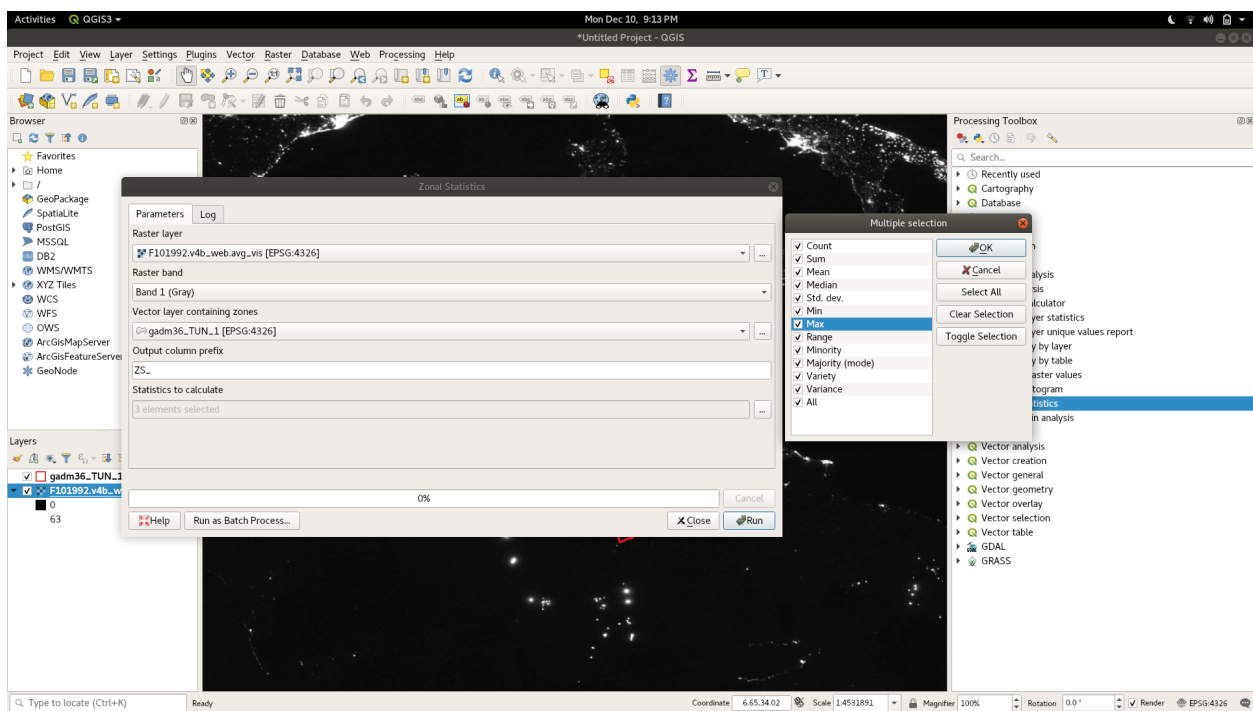Figure 1: QGIS Full Window

Figure 2: QGIS Raster popup

Figure 3: QGIS Zonalstat