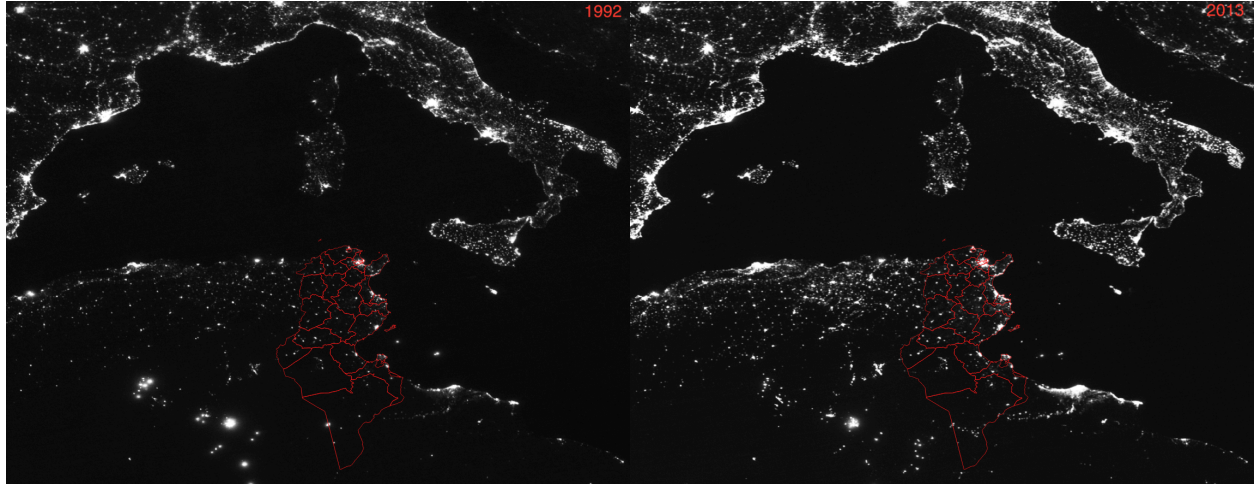


Report1

(1) Introduction

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



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

https://raw.githubusercontent.com/mariemayadi/data/master/tun_lum_governorate_93_13_tidy.csv

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 filling 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 organism 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.



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 question. As a team, we then were able to modify, integrate and create a story line around 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

(3) Analysis of data quality

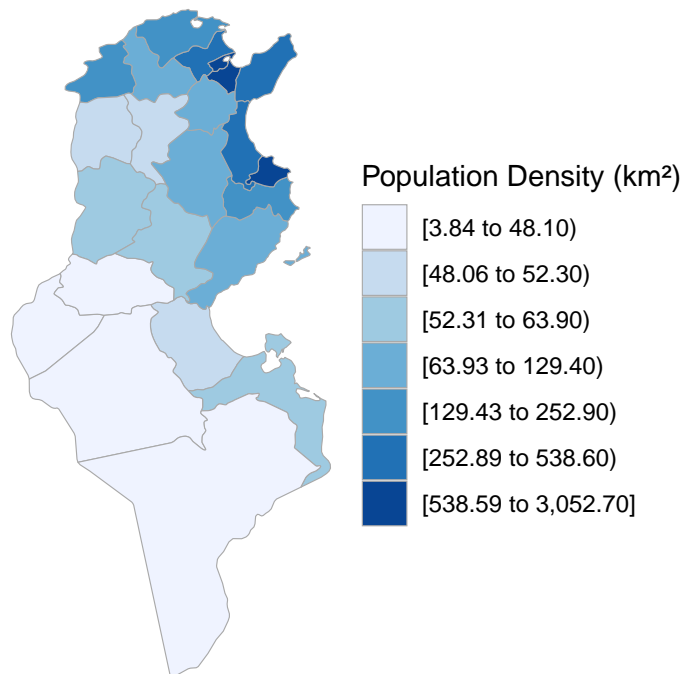
Some of the variable names are unclear, probably due to shoddy translation (i.e. what is the diff between 'Number of households having drinking water from the public source or source of water association' and 'Number of households having drinking water from the other public or private source'? - translation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

(4) Main analysis (Exploratory Data Analysis)

Extensive data processing

(4.a) A First Look at Tunisia

Tunisia's Population Density by Governorate
Year: 2014

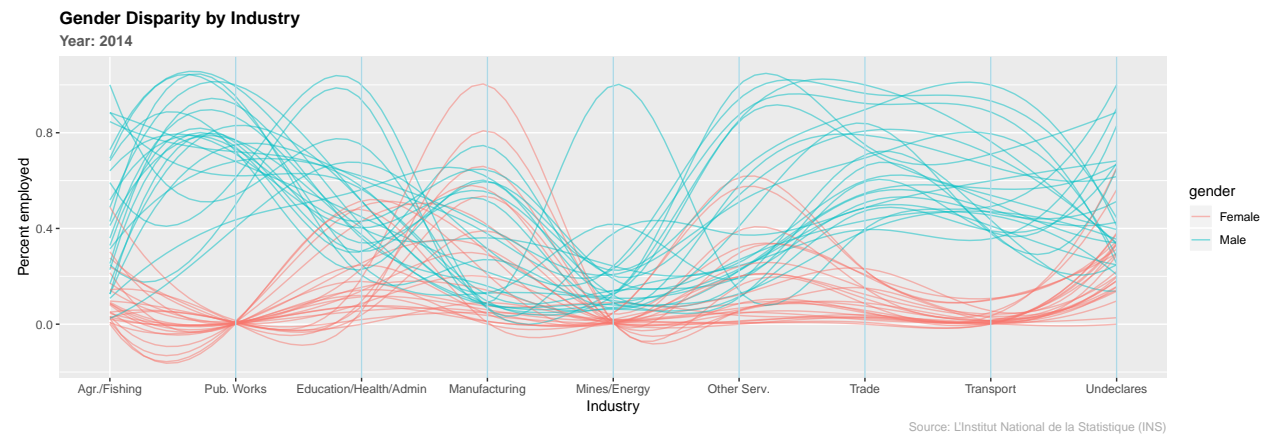


Source: L'Institut National de la Statistique (INS)

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 indicators 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 choroplethAdmin1 package, so all choropleth plots of Tunisia on the governorate level will show Ariana and Manouba as a combined governorate.

(4.b) Population breakdown by Industry and Gender

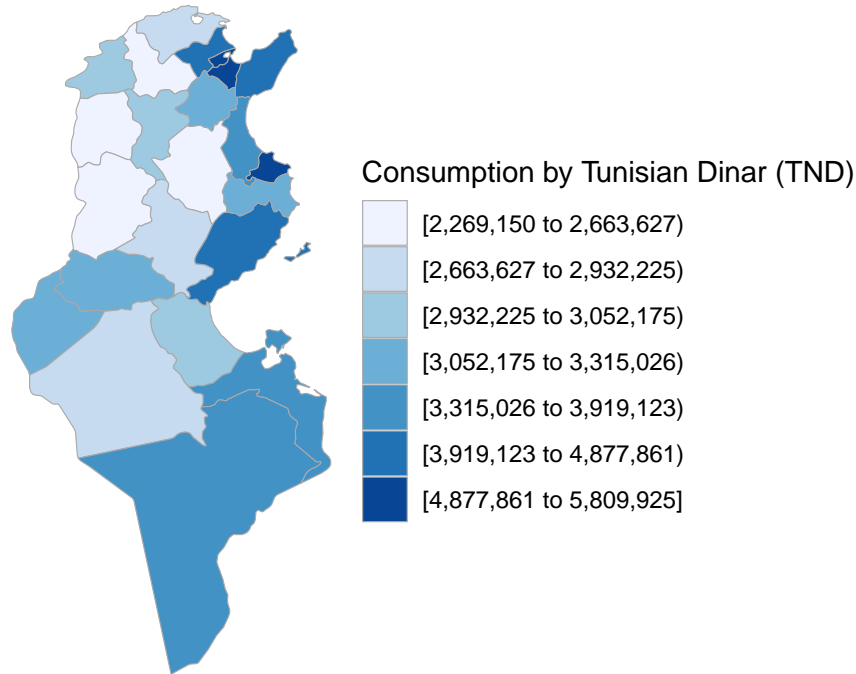


This plot shows one line per governorate per gender for the year 2014, with the Y value representing the fraction of total persons fifteen years or older that are employed in that industry. It allows us to observe disparities between employment sector trends between men and women, as well as which employment sectors tend to 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

Consumption by Governorate

Year: 2015



Source: 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.

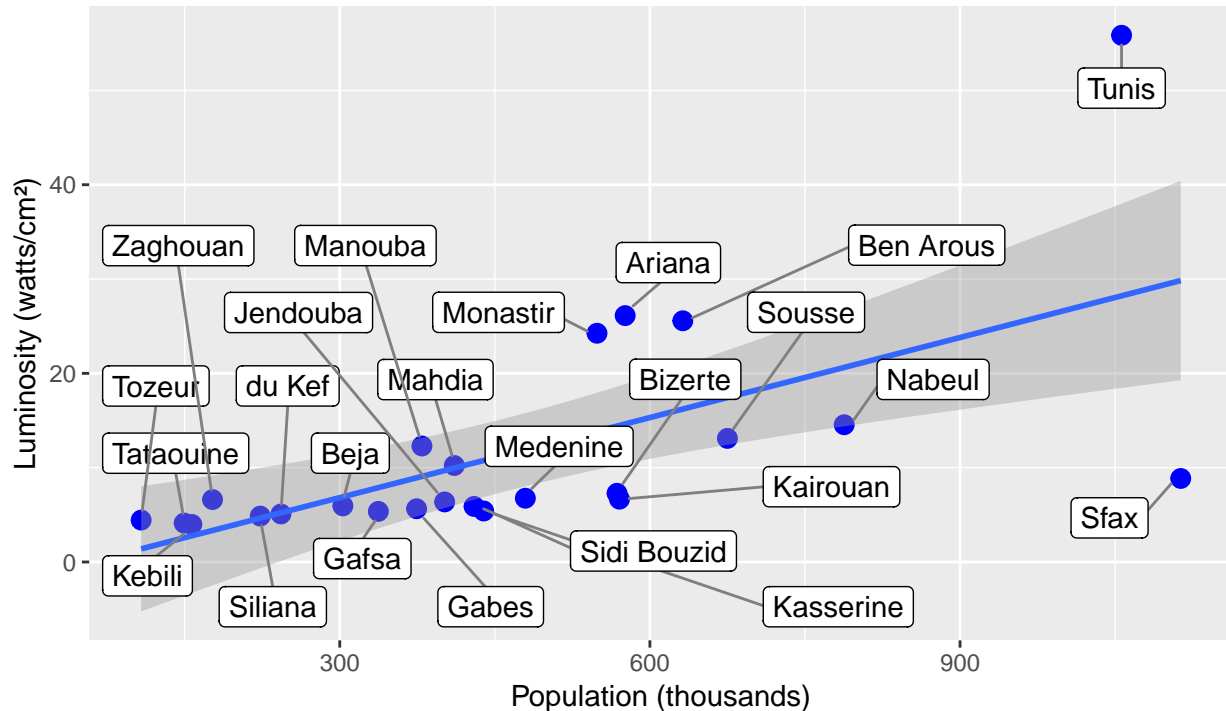
While the wealthy coastal governorates have the highest levels of consumption per capita, there are some subtle differences from the population density choropleth plot. Consumption is lowest in some of the inland northwest governorates, while the southern, sparsely populated governorates have relatively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

Luminosity Vs population by Governorates

Year: 2014



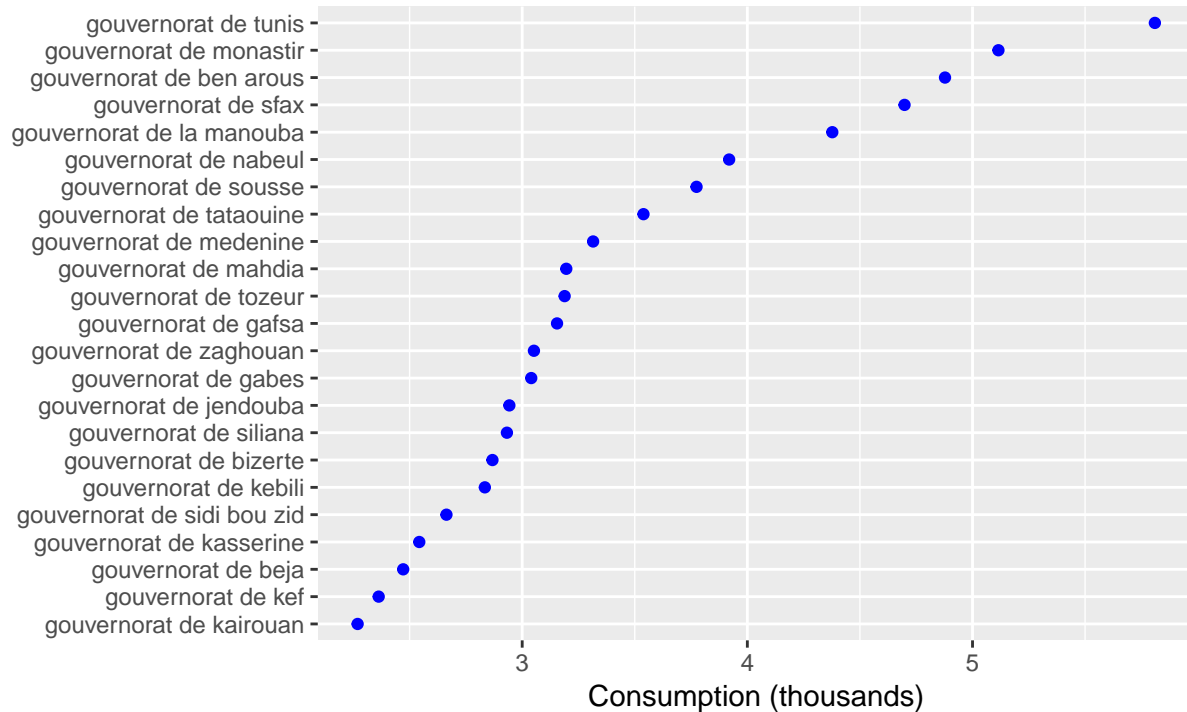
Source: National Oceanic and Atmospheric Administration (NOAA)

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 an 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.

Consumption by Governorate

Year: 2015



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

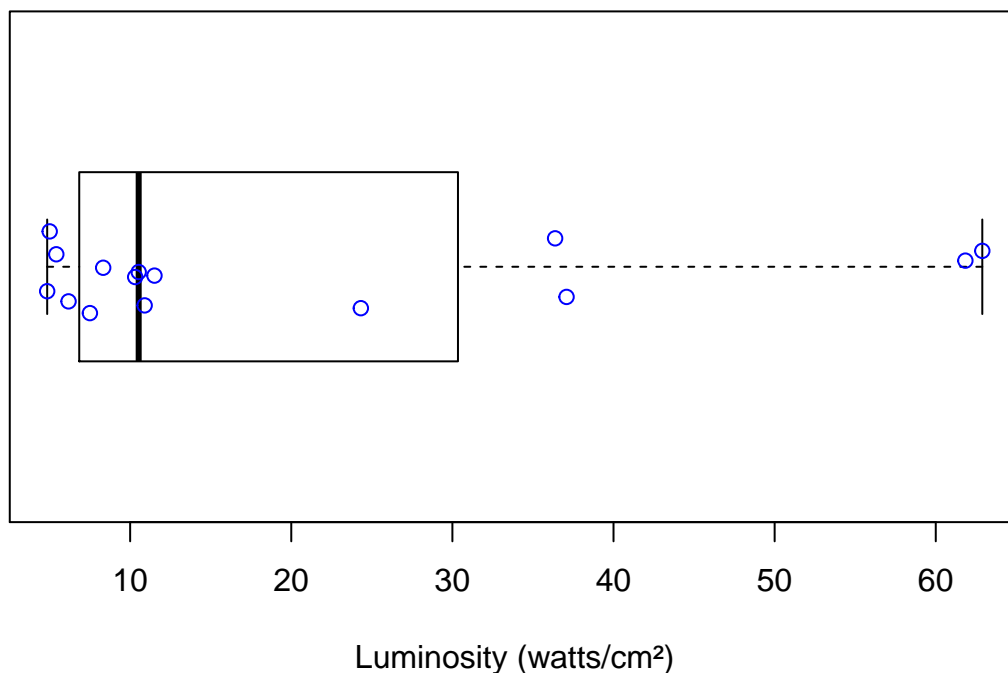
This is a dot plot of consumption per capita (see 4.c for a thorough definition of this variable) by governorate for the year 2015. Tunis has the highest consumption per capita in the country, while Monastir, Ben Arous, and Manouba, which are essentially three suburban regions close to Tunis geographically 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 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 dinar.

(4.e) A closer at outliers: Sfax

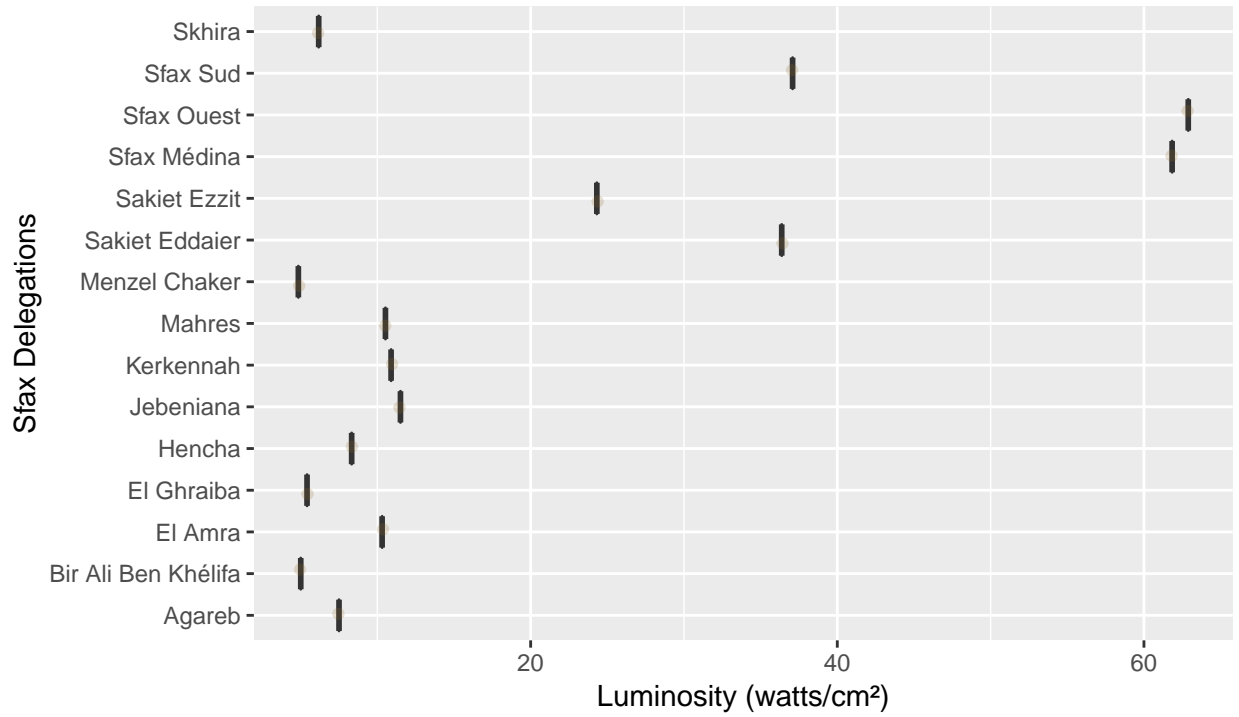
Luminosity of Sfax Delegations (Year: 2013)



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.

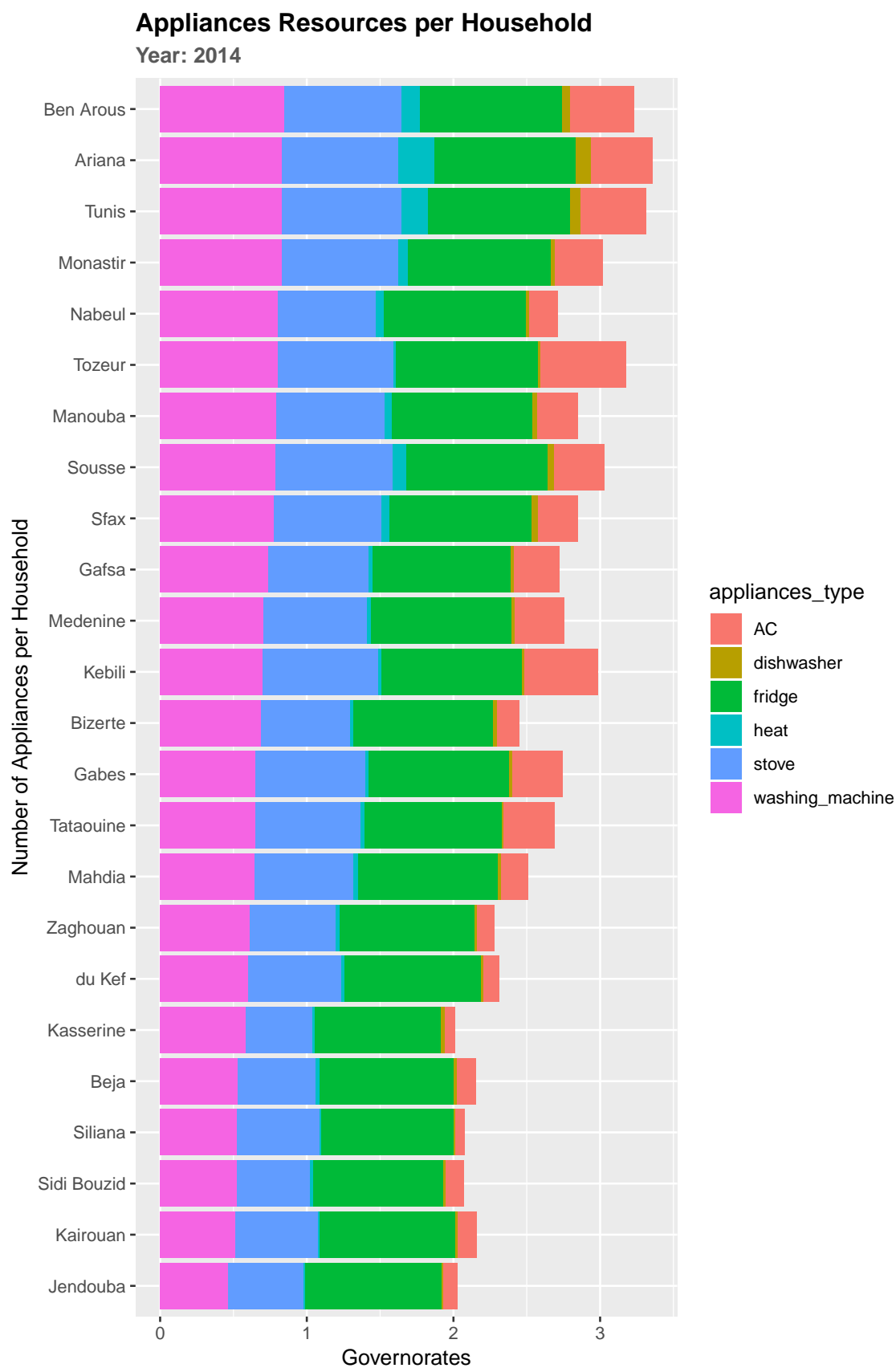
Which are the leading delegations within Sfax?

A closer look at luminosity by delegations within Sfax



Source: National Oceanic and Atmospheric Administration (NOAA)

(4.f) Exploring potential covariates of luminosity

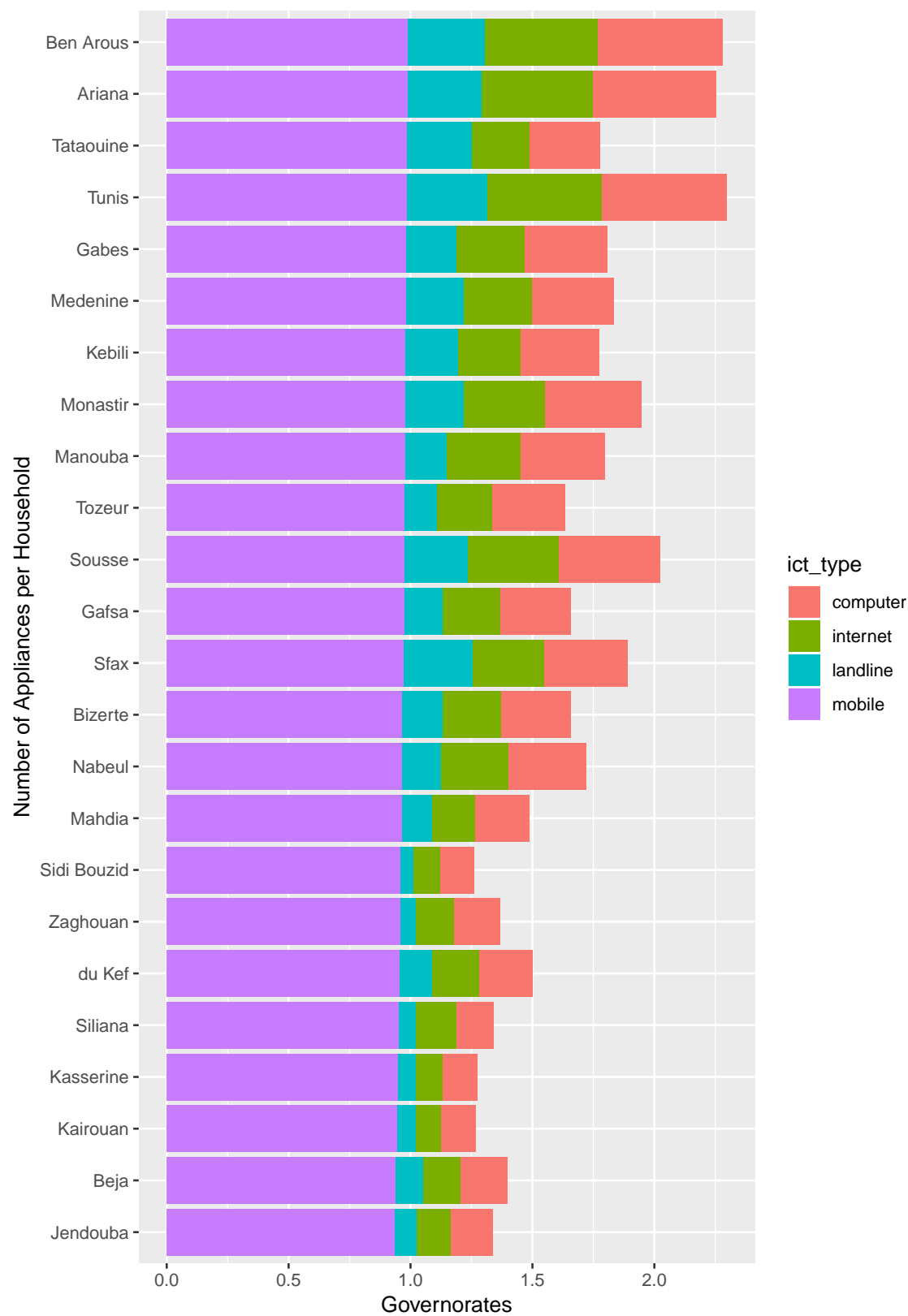


Source: L'Institut National de la Statistique (INS)

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.

Information and Communications Technology (ICT) Resources per Governorate

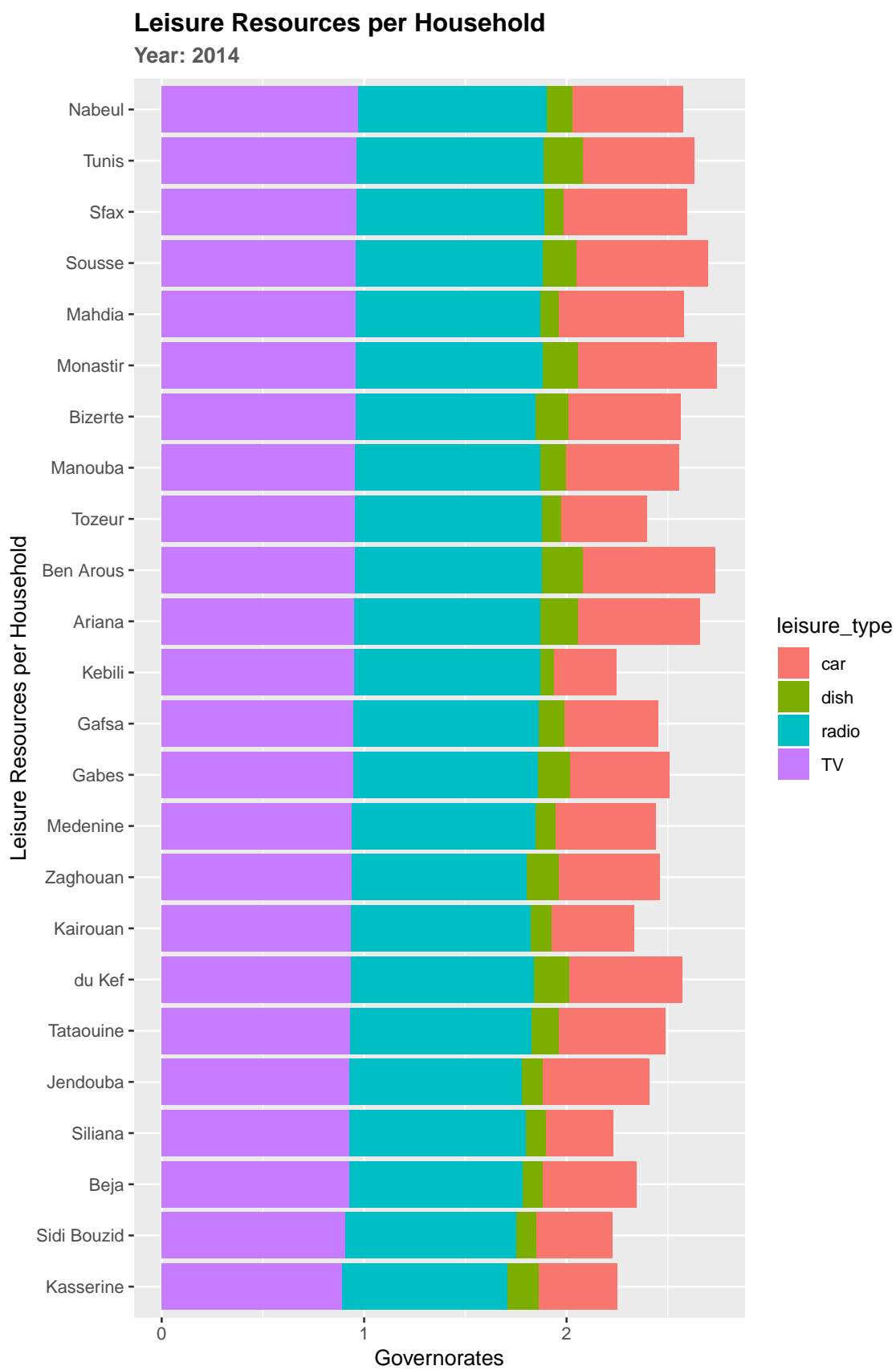
Year: 2014



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

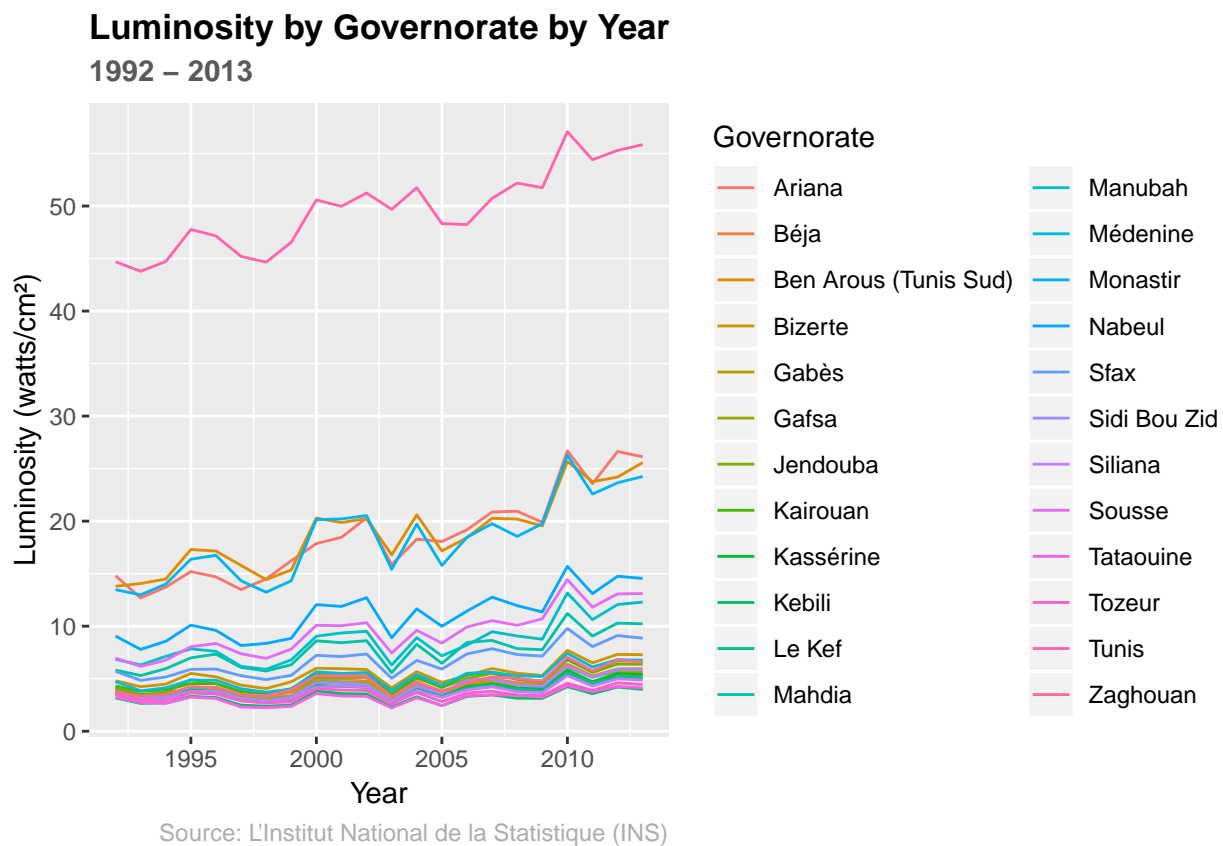


Source: L'Institut National de la Statistique (INS)

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 glean 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 consistently measured throughout the years in Tunisia.

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



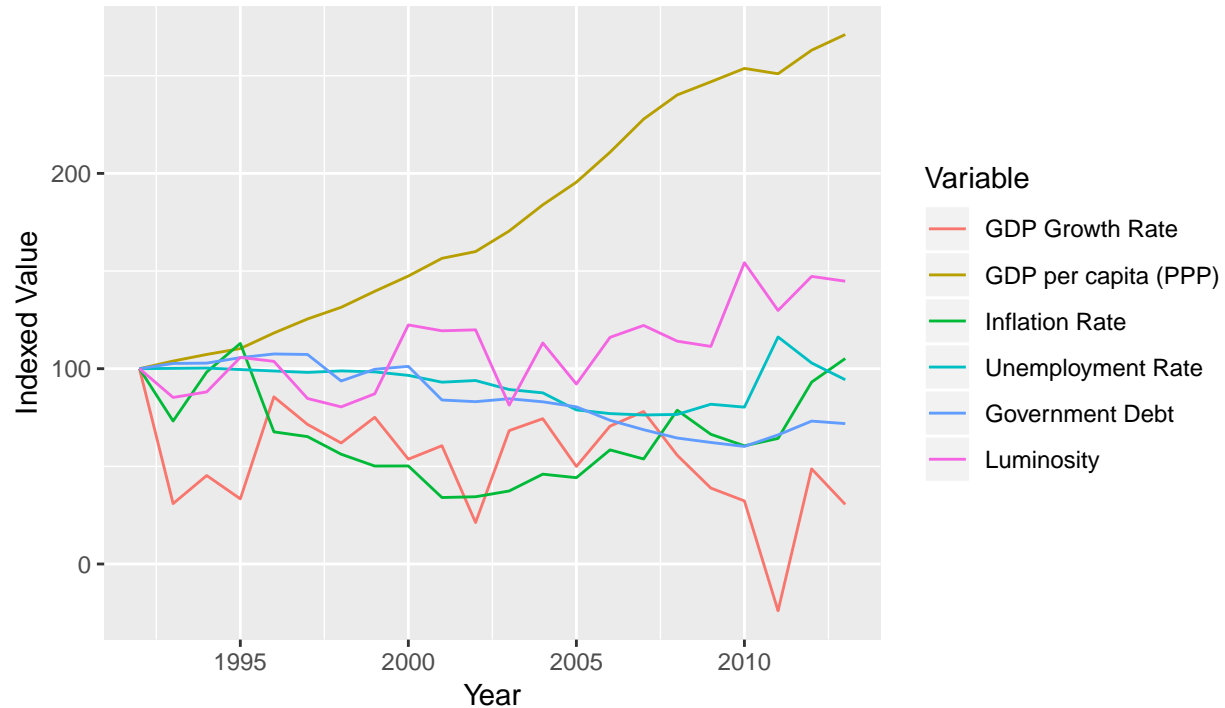
This plot of luminosity by governorate by year shows that Tunis has had a significantly higher level of luminosity than the other governorates. The three governorates in the middle of the graph (Ariana, Ben Arous, and Monastir) are suburban extensions of Tunis. The rest of the governorates are bunched 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.

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.

The variable 'avg_lum' 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.

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

	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

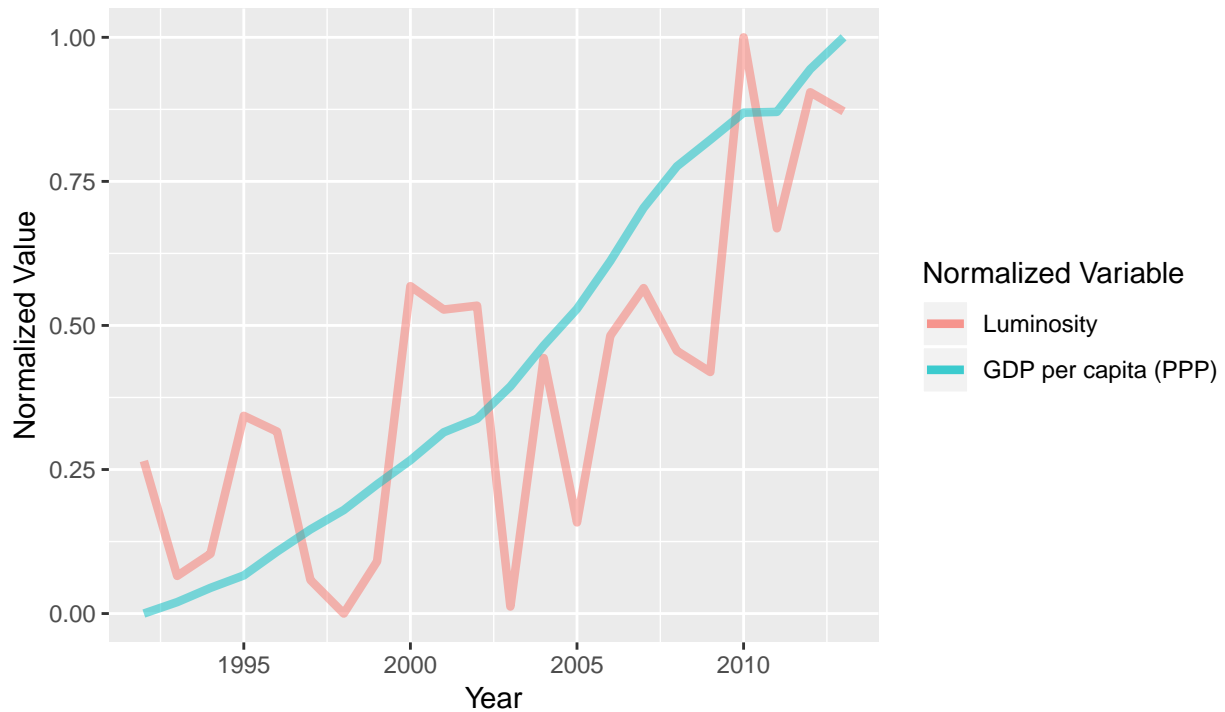
```

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.

Luminosity vs. GDP per capita (PPP) in Tunisia (Normalized)

1992 – 2013

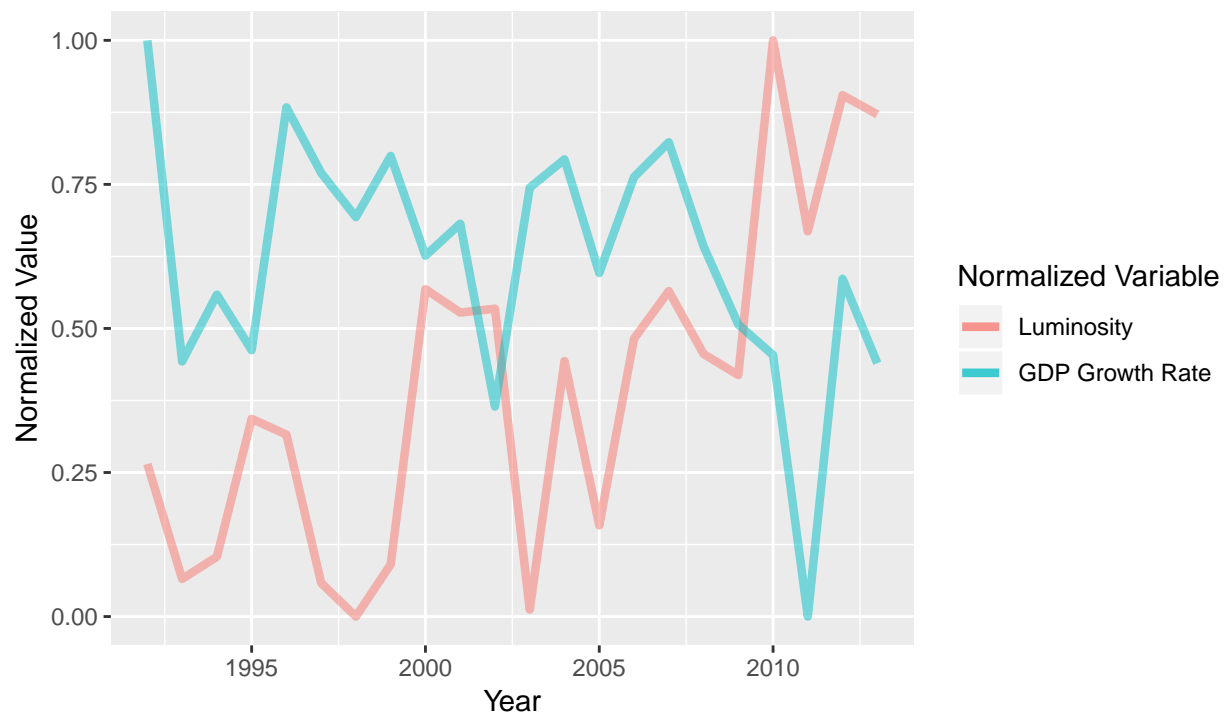


s: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database

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.

Luminosity vs. GDP Growth Rate in Tunisia (Normalized)

1992 – 2013

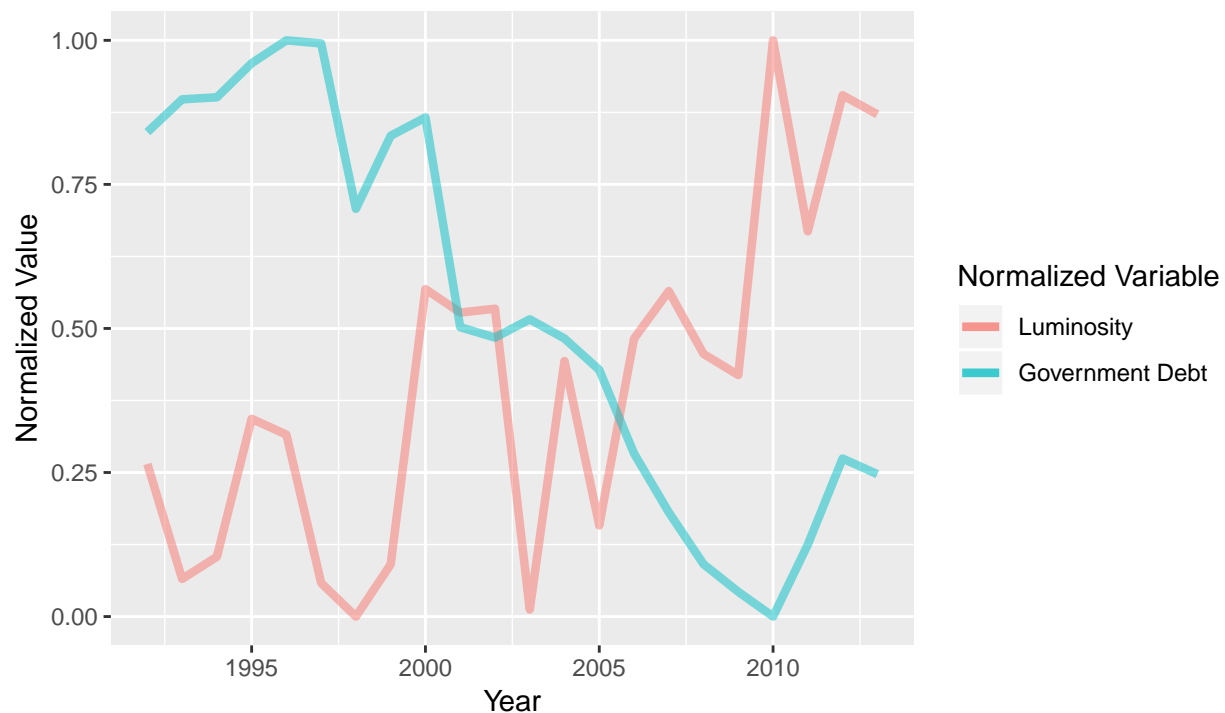


ources: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database

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.

Luminosity vs. Government Debt in Tunisia (Normalized)

1992 – 2013



ources: L'Institut National de la Statistique (INS), IMF World Economic Outlook Database

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 (Presentation-style)

(6) Interactive component

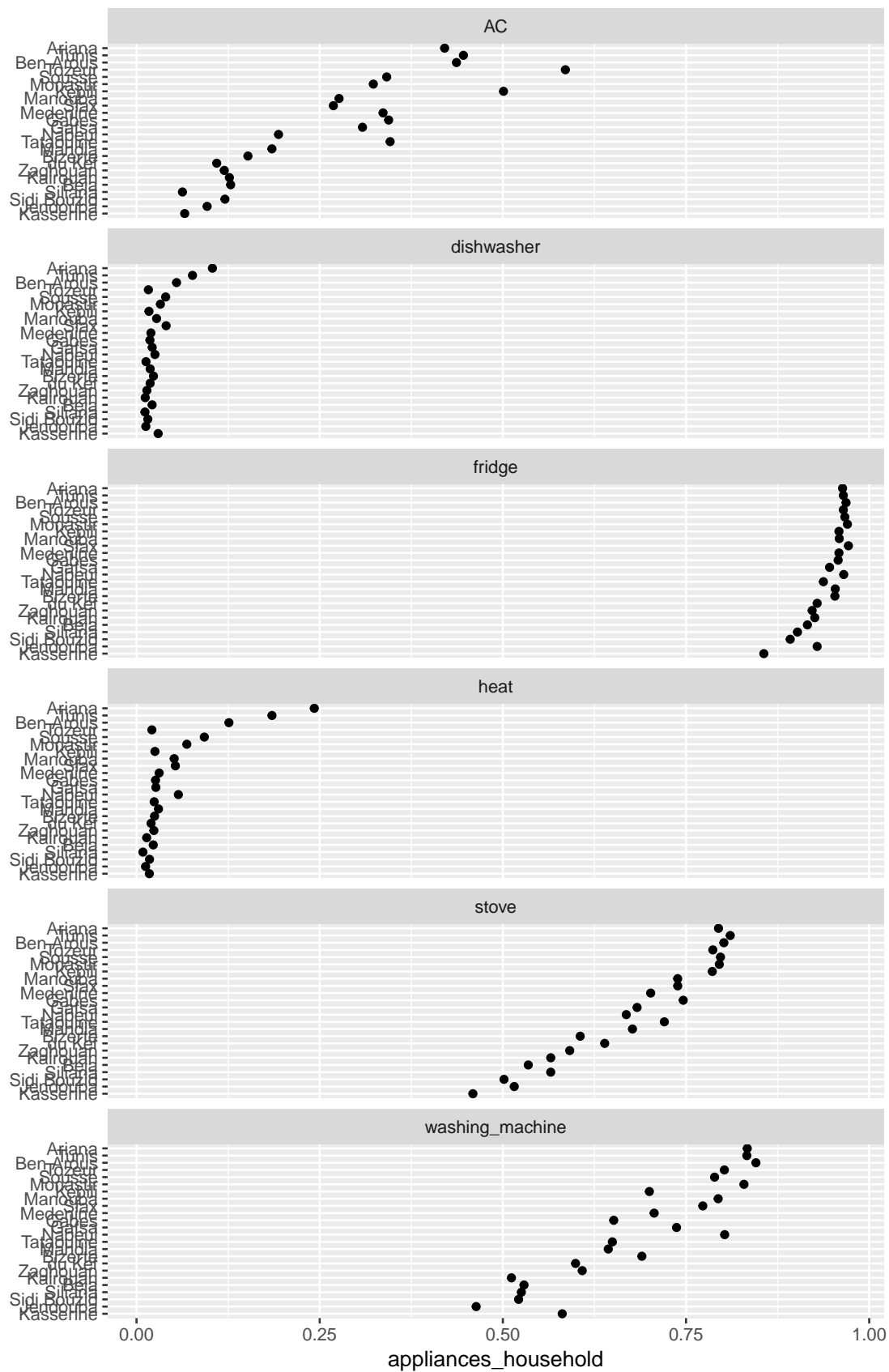
(7) Conclusion

- Interested in looking at 2011 (year of the jasmine revolution) since the GDP went drastically down (Maybe less economic activity? Look into luminosity?)
- 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: choropleth naming for governorates is very specific and does not support french accent". Two regions are missing.
- Bin side - looked into customizing it but very time consuming

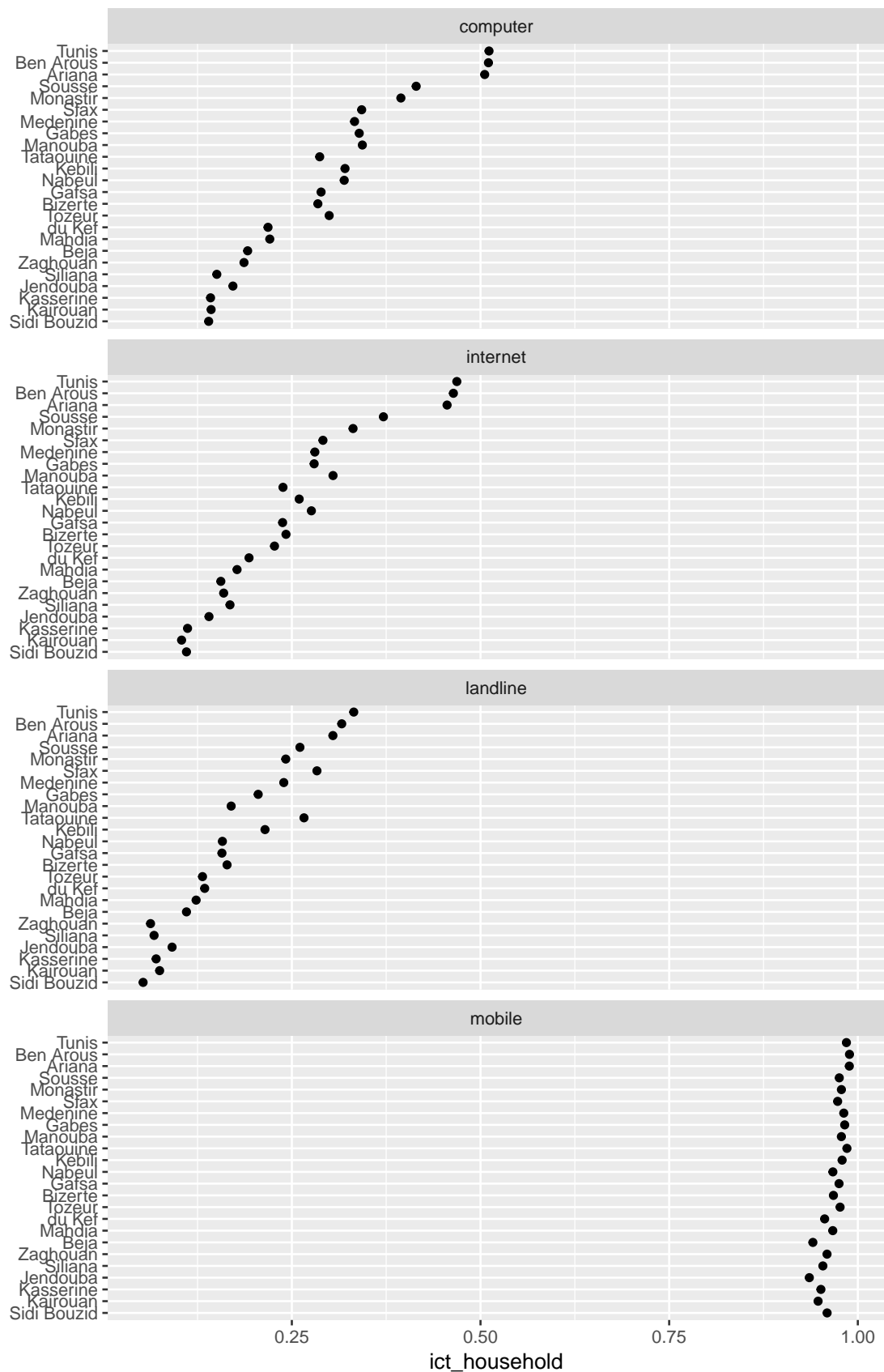
Appendix

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.

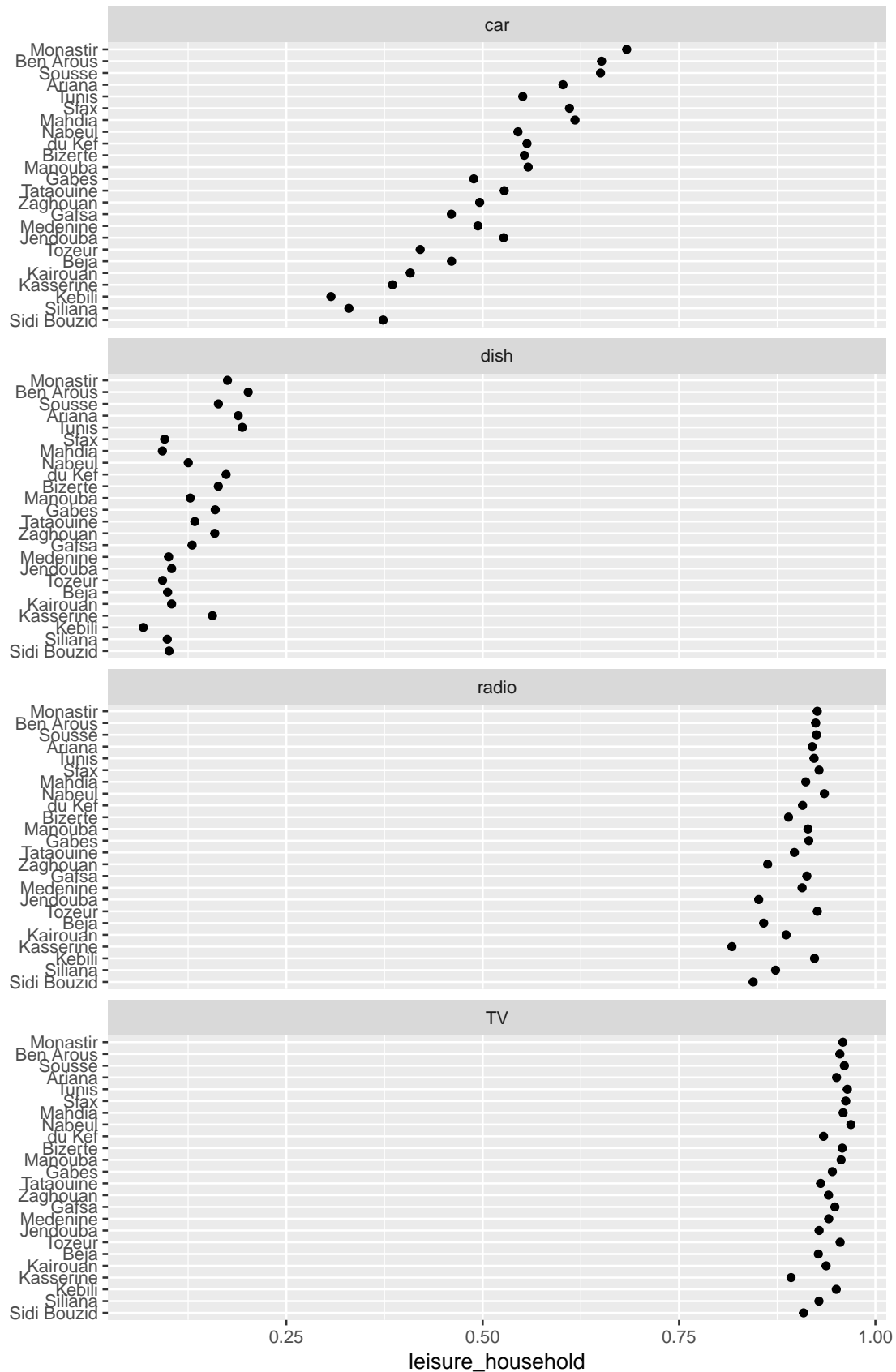
appliances Resources per Household

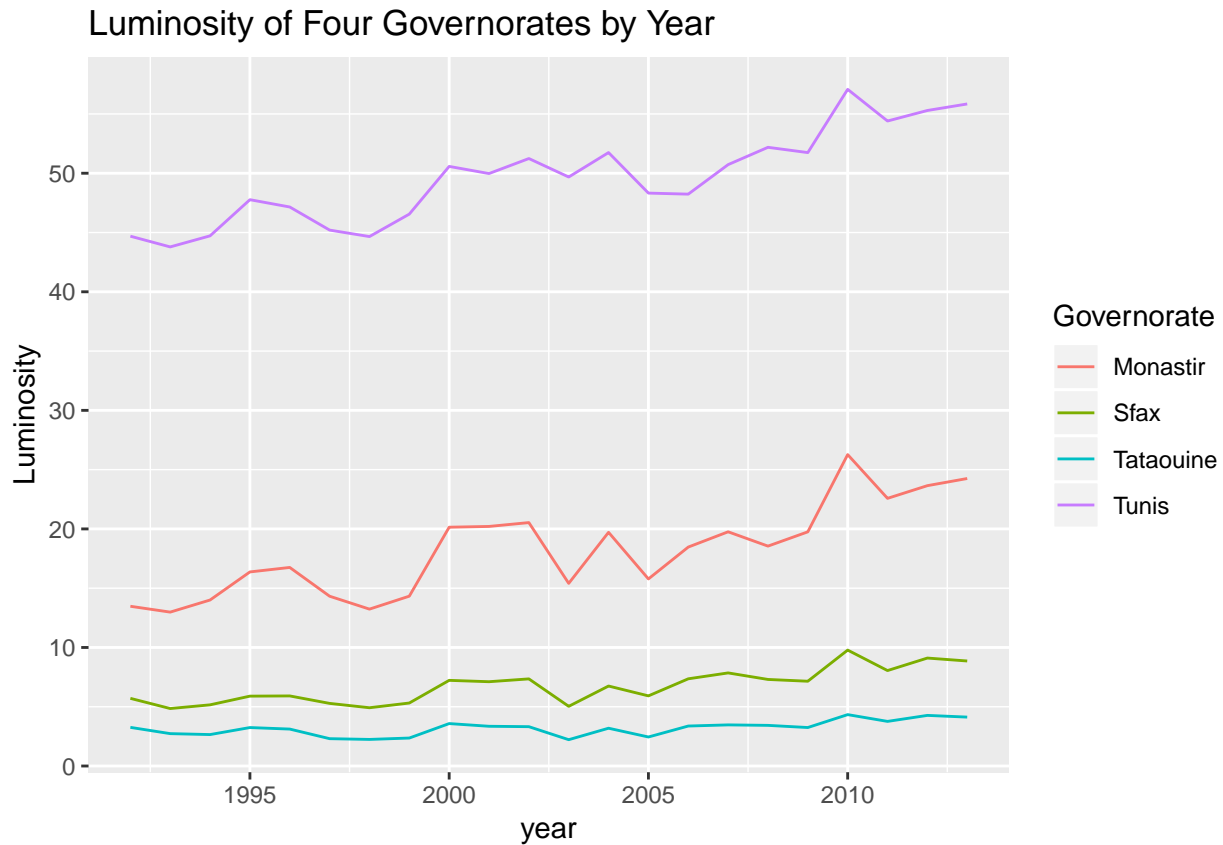


internet Resources per Household



Leisure Resources per Household





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 Tataouine 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.