Report1

(1) Introduction

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

(2) Description of data

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

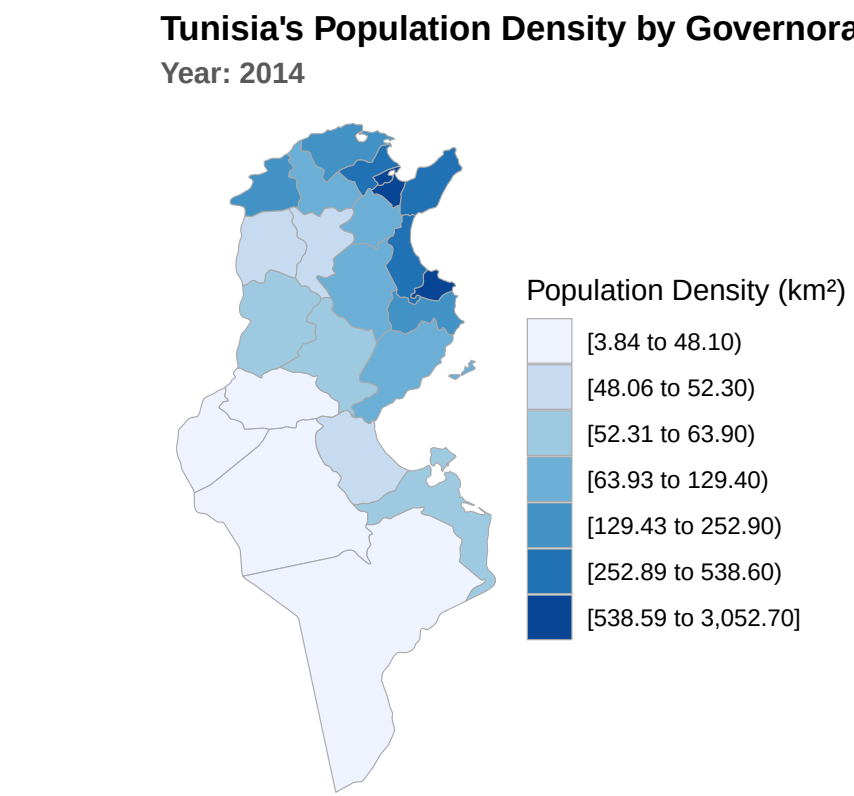
(3) Analysis of data quality

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

(4) Main analysis (Exploratory Data Analysis)

Extensive data processing

(4.a) A First Look at Tunisia

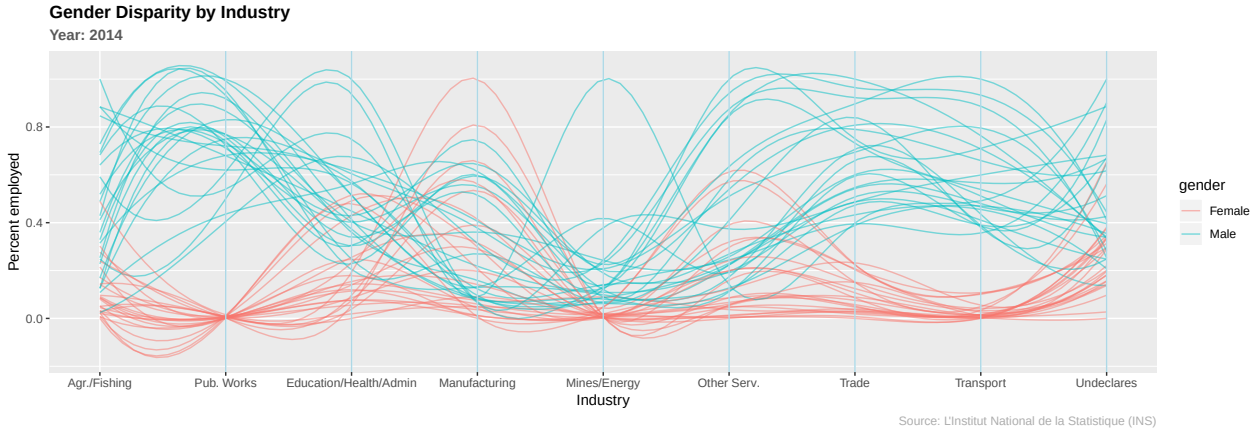


Source: L'Institut National de la Statistique (INS)

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

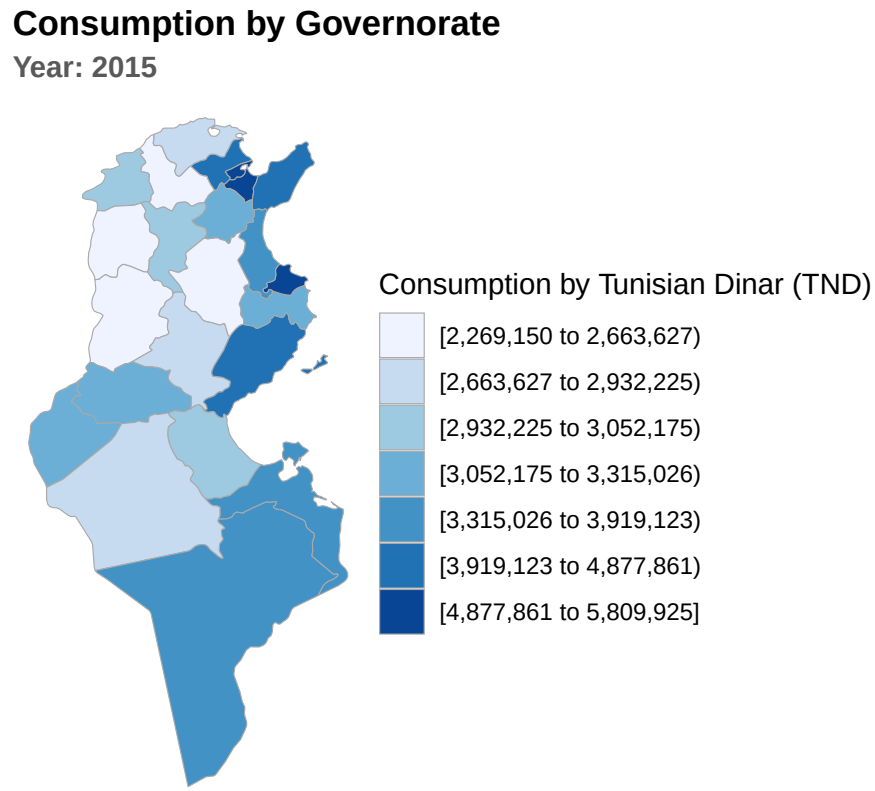


Source: L'Institut National de la Statistique (INS)

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

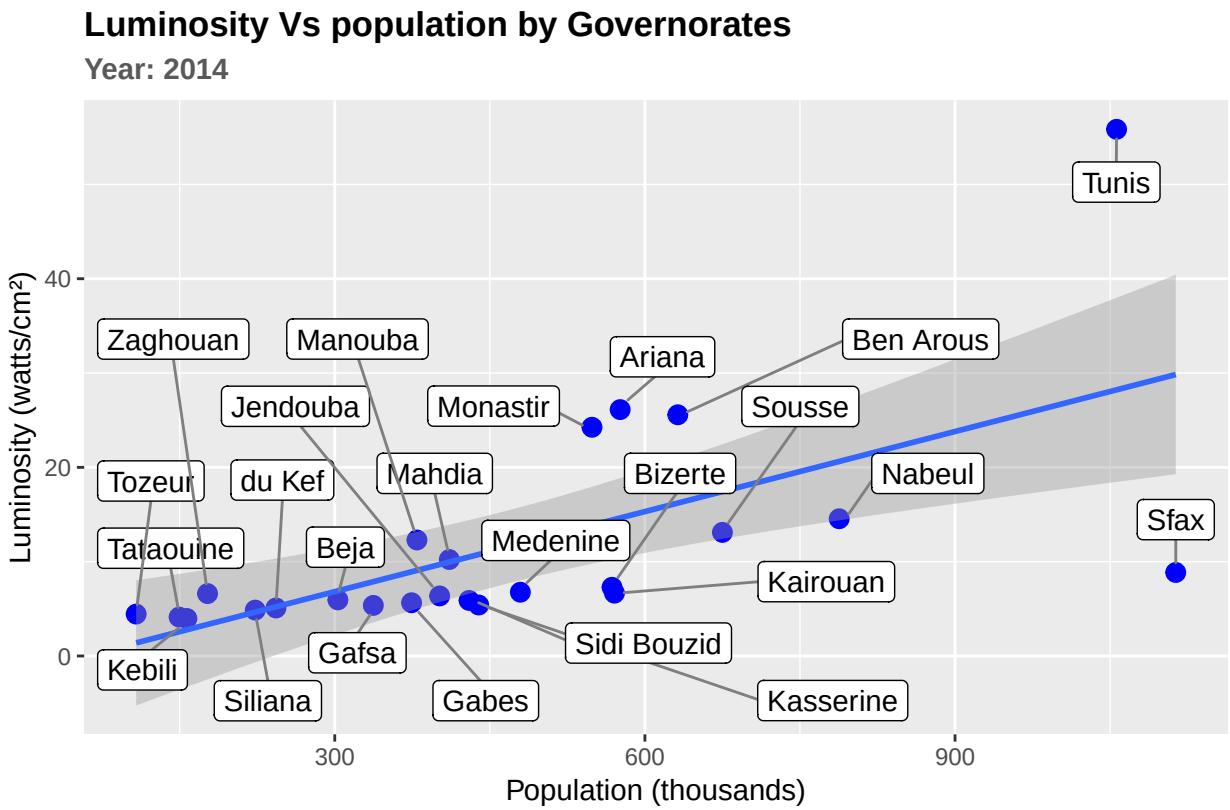
TODO: - add after map bar plot or cleveland dot plot to show ranking of governorates by consumption



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

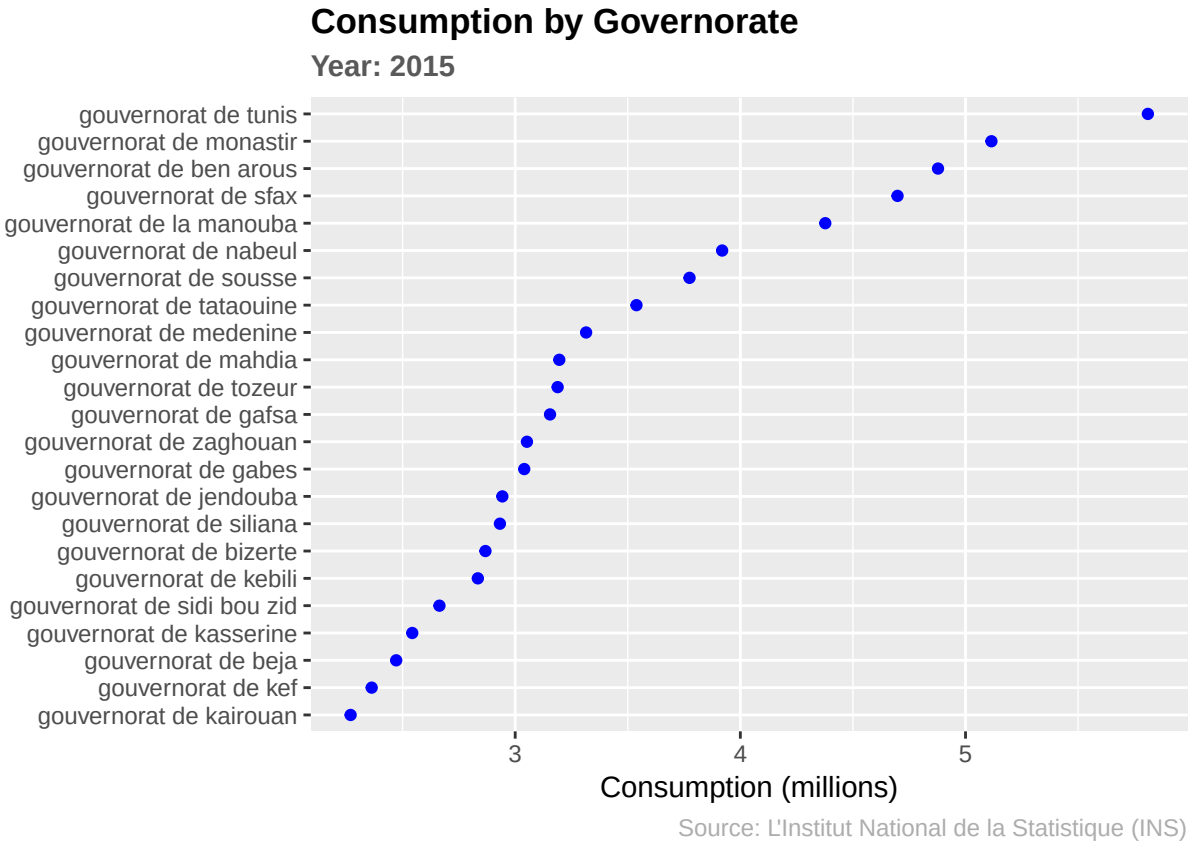
Comments: - definie consumption - can include excat categories that made up total - per capital - first comma from the right is delignating cents - show top 5 - focus on sfax

(4.d) Luminosity trends across Tunisia



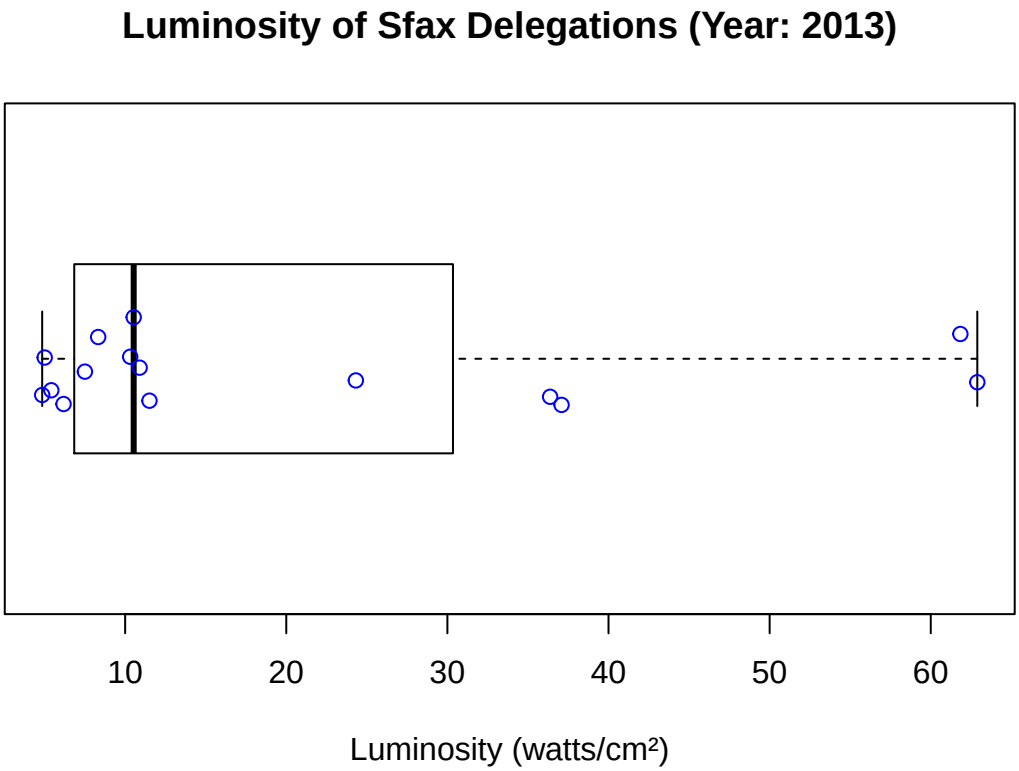
This scatter plot of mean luminosity (in watts/cm2) against population (in thousands) by governorate shows some interesting trends, especially for the two high-population governorates of Tunis and Sfax to the right of the graph. Overall with the loess smoother, we can see that as expected there is a 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.

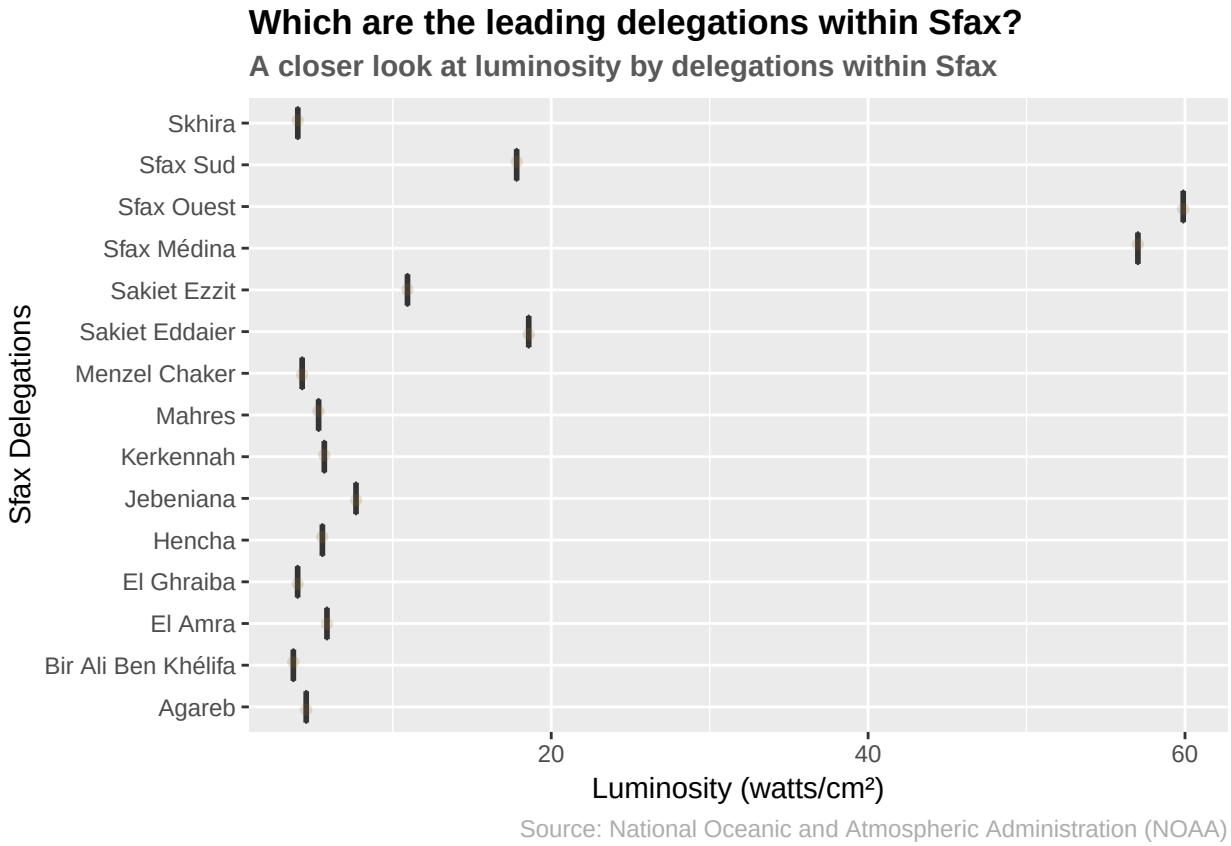
In contrast, Sfax (population 1,113,496) is sharply below the trendline. While Sfax is even more populous than Tunis, it has a significantly lower level of mean luminosity relative to population than the rest of Tunisia. This is a rather bizarre finding that merits further exploration. We decided to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



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

(4.e) A closer at outliers: Sfax

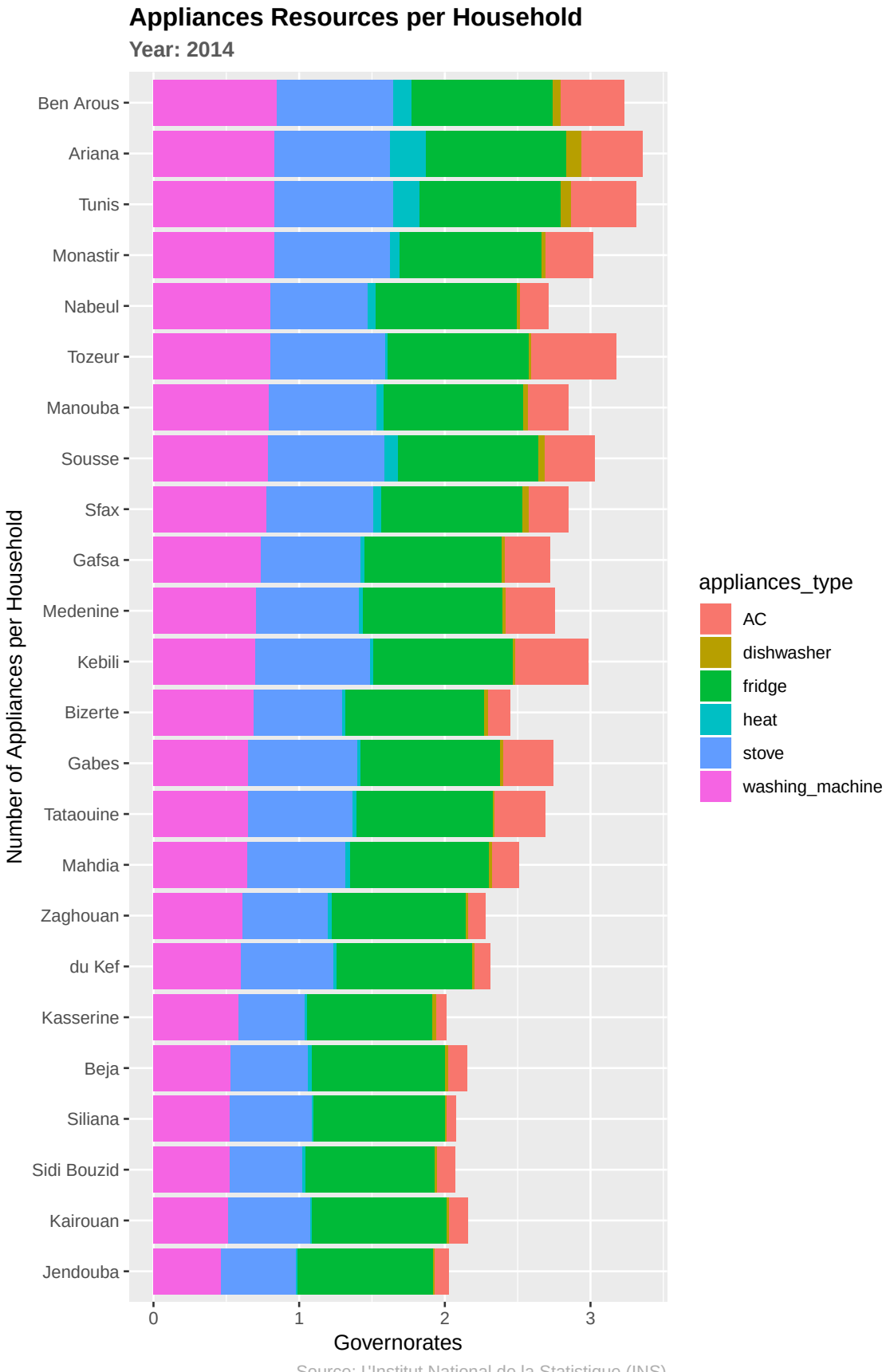




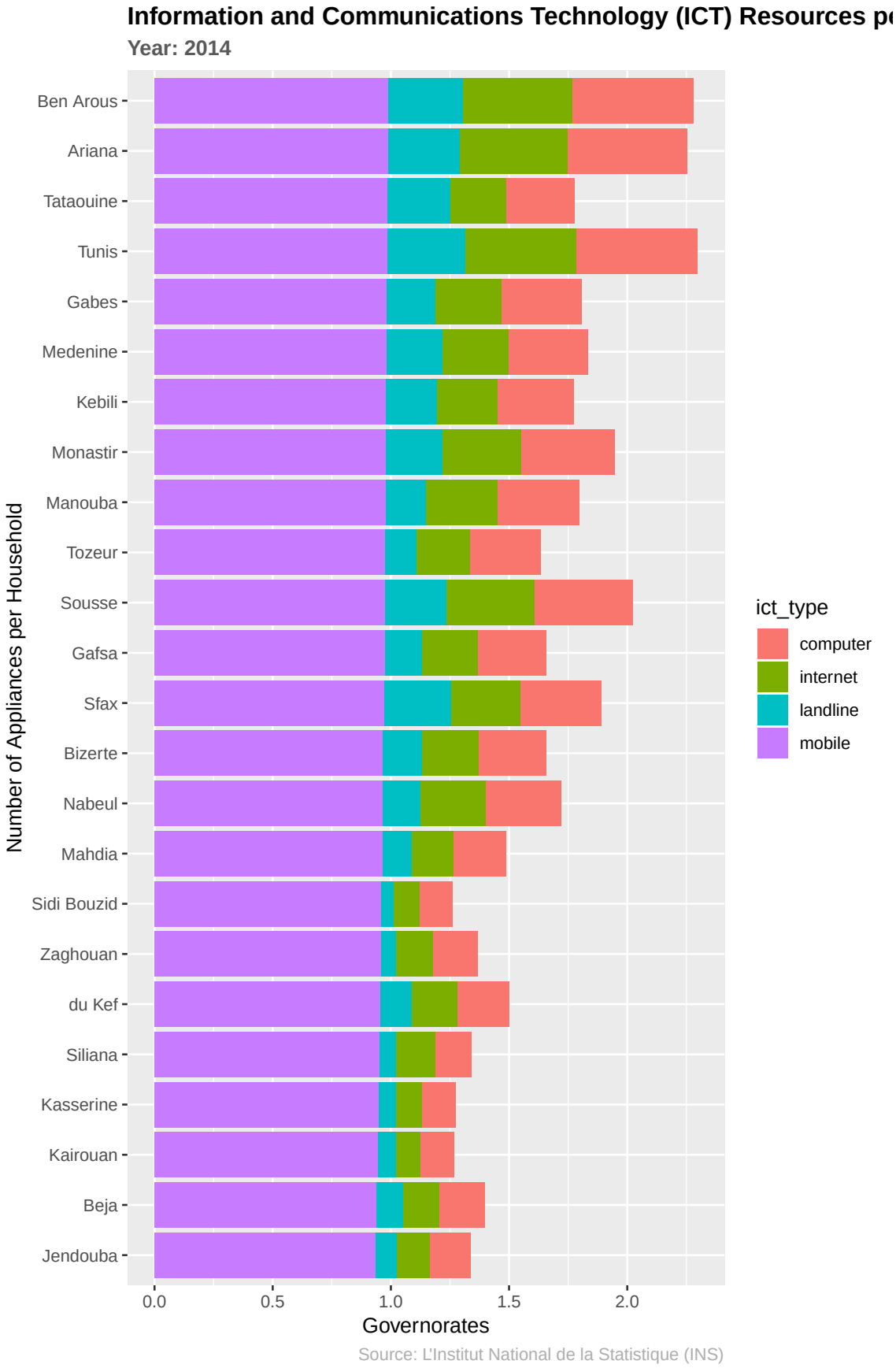
Comments: - Clear unbalance between delegations within the Sfax governorates. The outliers are: xxx and YYY.

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

(4.f) Exploring potential covariates of luminosity

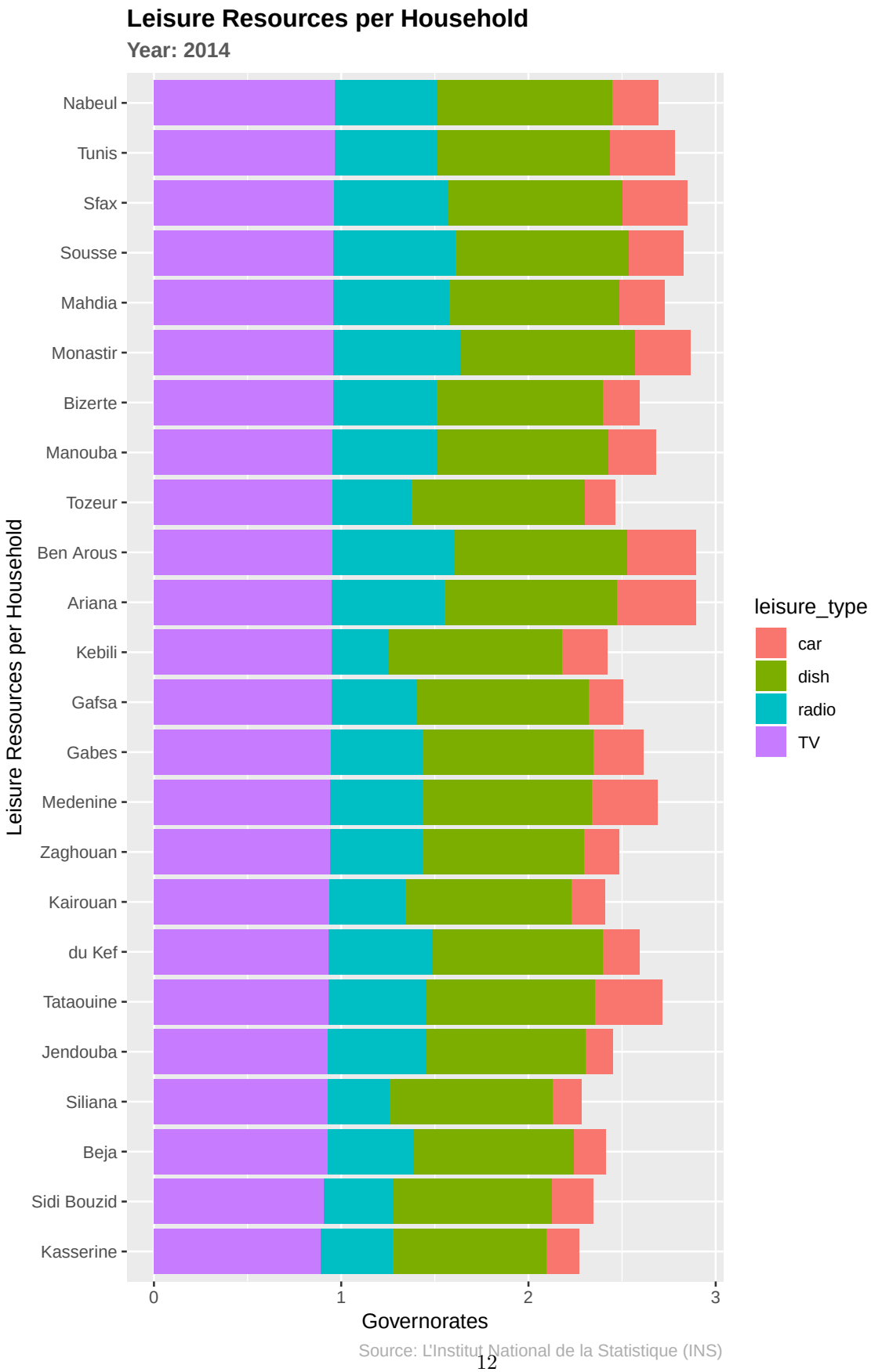


This plot compares the numbers of functional appliances per household by governorate. Appliances examined are air conditioners, dishwashers, refrigerators, heaters, stoves, and washing machines. Refrigerators are the most ubiquitous appliance, with the average per household being nearly 1 across the 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 pe
Year: 2014
Ben Arous
Ariana
Tataouine
Tunis
Gabes
Medenine
Kebili
Monastir
Manouba
Tozeur
Sousse
Gafsa
Sfax
Bizerte
Nabeul
Mahdia
Sidi Bouzid
Zaghouan
du Kef
Siliana
Kasserine
Kairouan
Beja
Jendouba
Number of Appliances per Household
0.0
0.5
1.0
1.5
2.0
Governorates
ict_type
computer
internet
landline
mobile
Source: L'Institut National de la Statistique (INS)

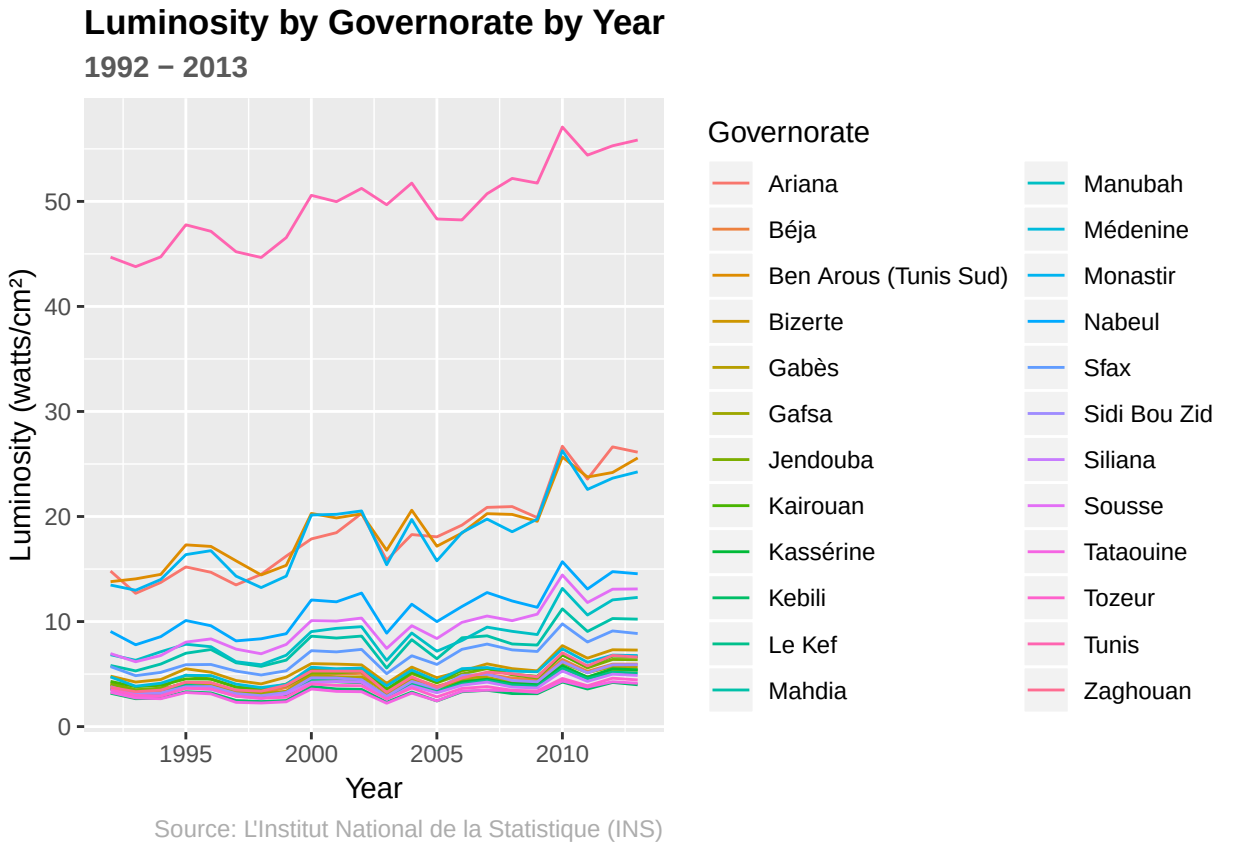
This plot compares the number of Information and Communication Technologies (ICT) resources per household by governorate. ICTs examined are mobile phones, landline phones, internet connections, and computers. We quickly see that mobile phones have a similar role in this plot to refrigerators in the previous 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



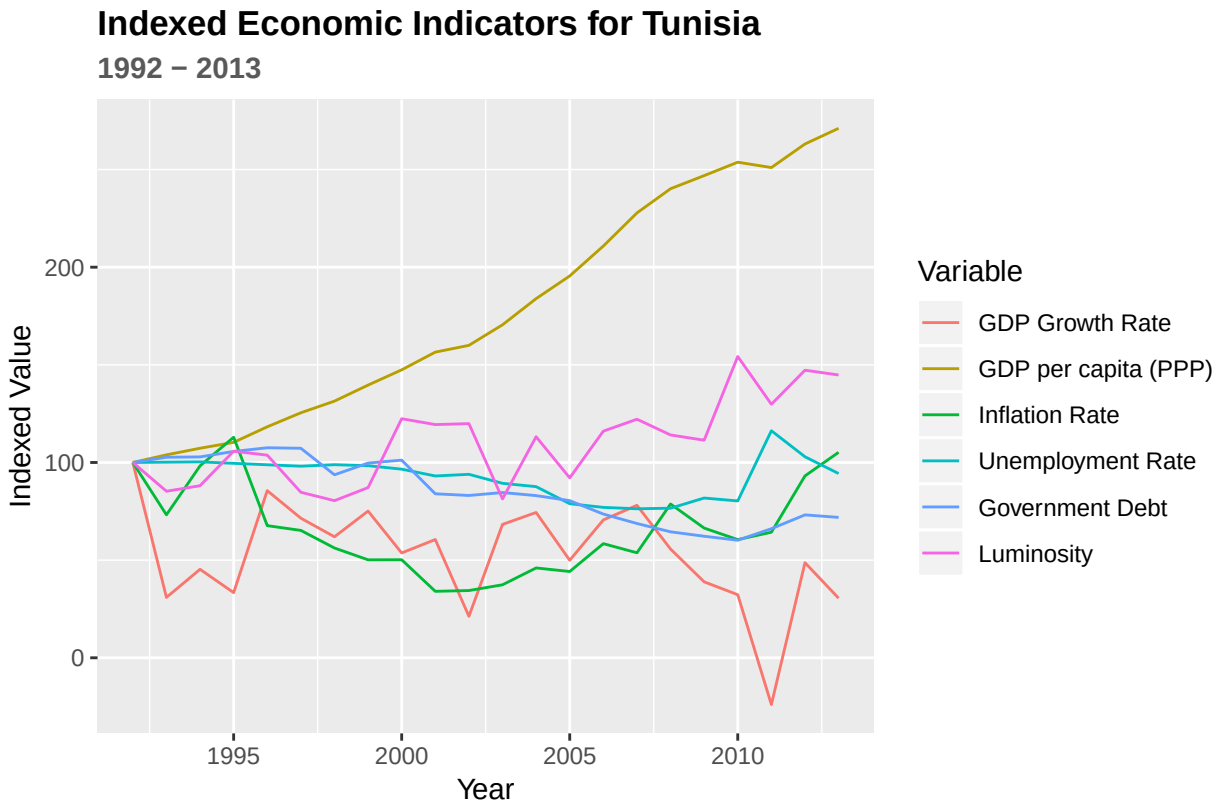
This plot compares the number of leisure resources per household by governorate. Leisure resources examined are televisions, radios, satellite dishes (for television), and cars. Nearly every household has at least one television and radio in all governorates. Satellite dishes are relatively rare but 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.



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.



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.

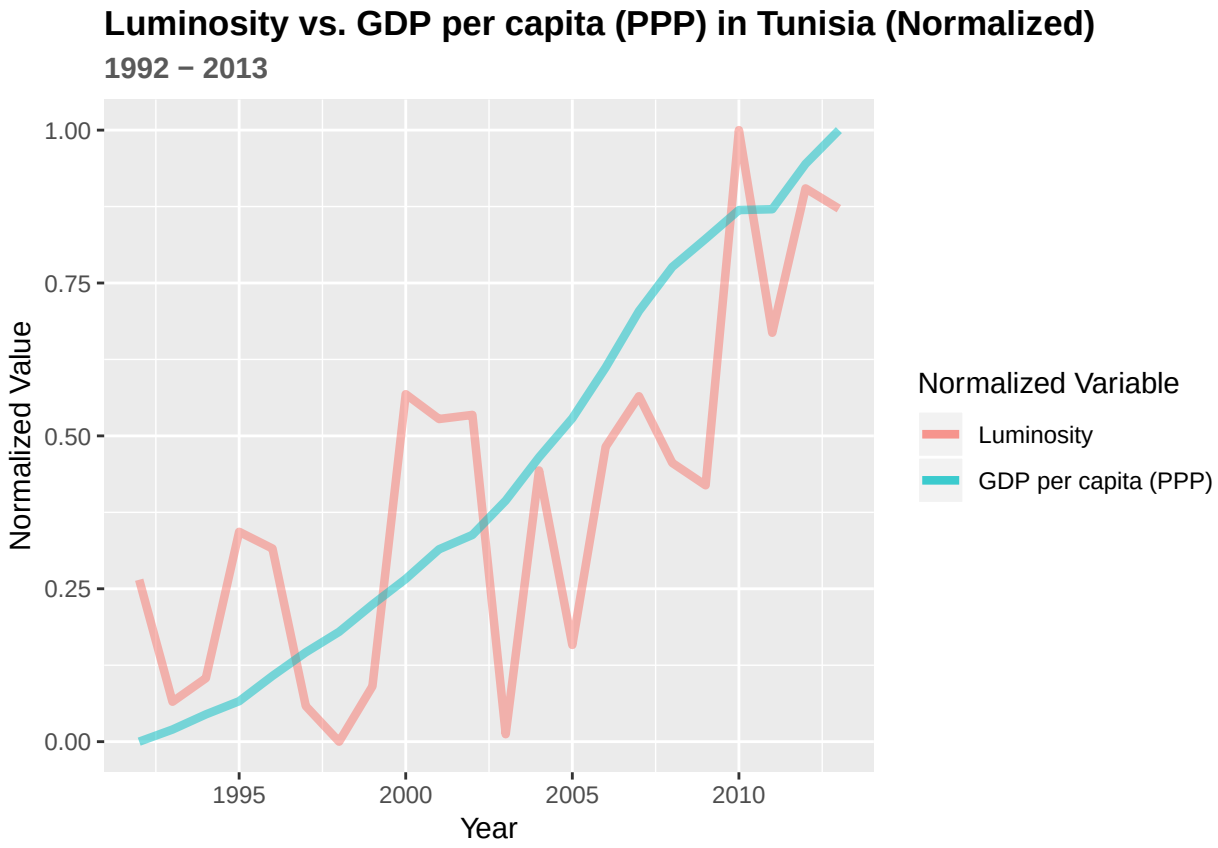
```
## Warning in cor(tunisia): the standard deviation is zero
```

```
##                      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
```

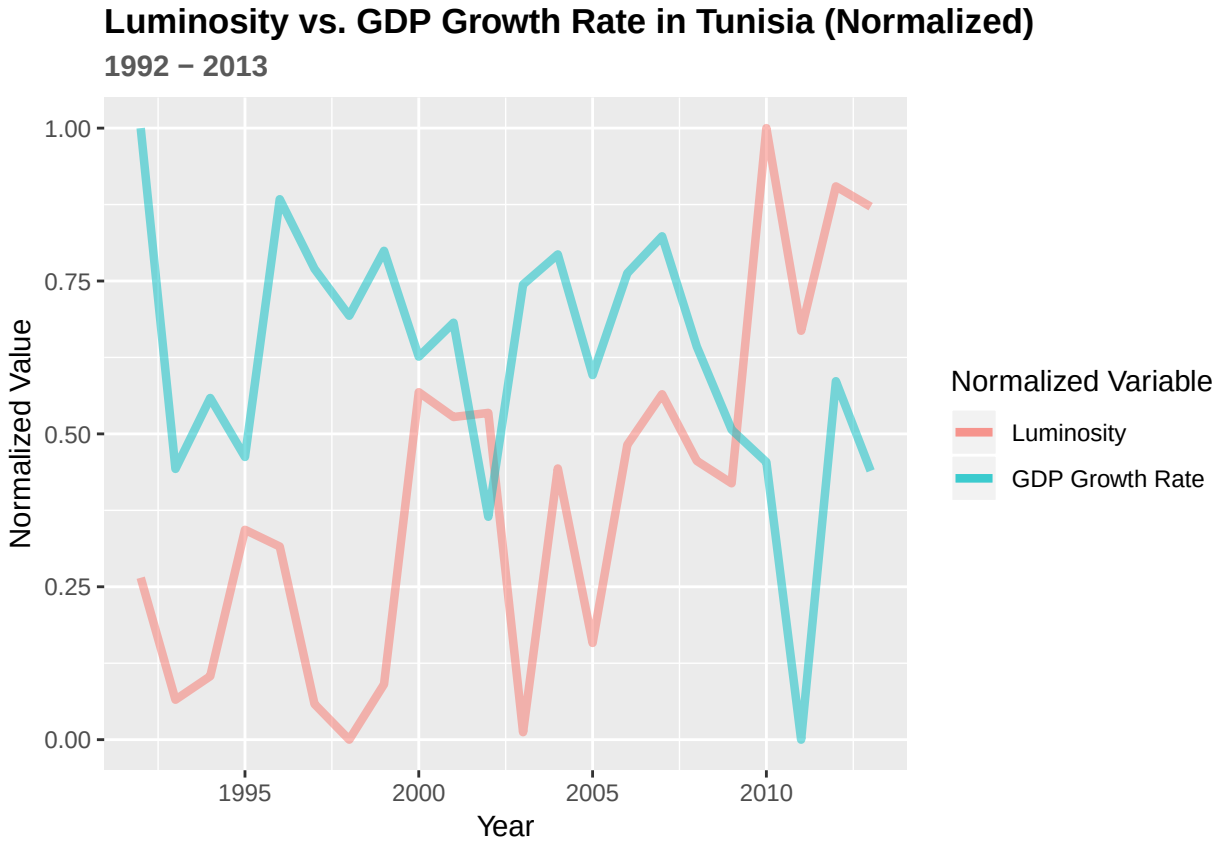
```
## inflation         -0.1098110    NA  0.1436795 -0.1029147 -0.008166219
## unemployment_rate -0.3512514    NA -0.1185138 -0.3242774 -0.355587641
## govt_debt         -0.9141032    NA -0.6580060  0.3674089 -0.927997907
## avg_lum            0.7329709    NA  1.0000000 -0.3818728  0.751756723
##                   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.

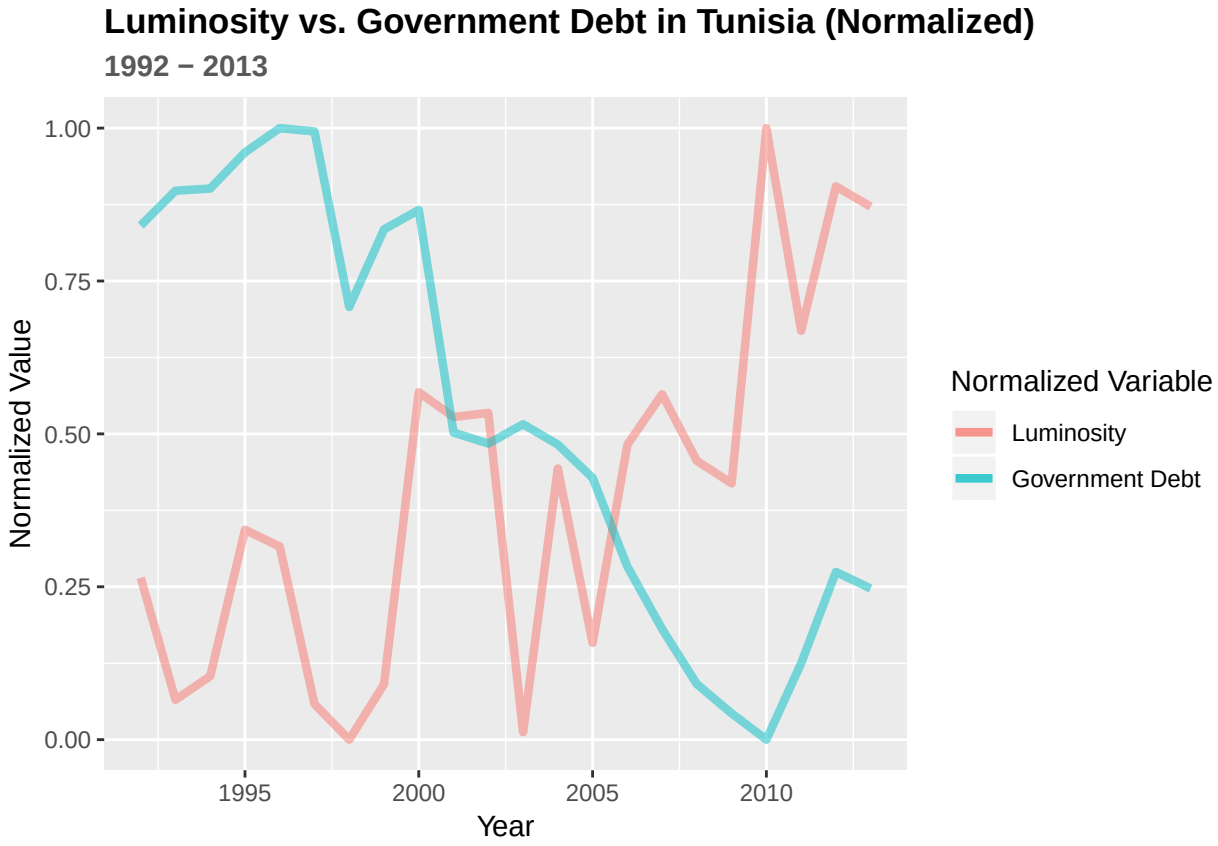
Of particular interest is that the relationship between luminosity and government debt is very strong in the negative direction (-0.66). A possible explanation for this is that higher levels of government debt are associated with the economy not doing well, and thus there is less luminosity to be observed.



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.



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.



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)

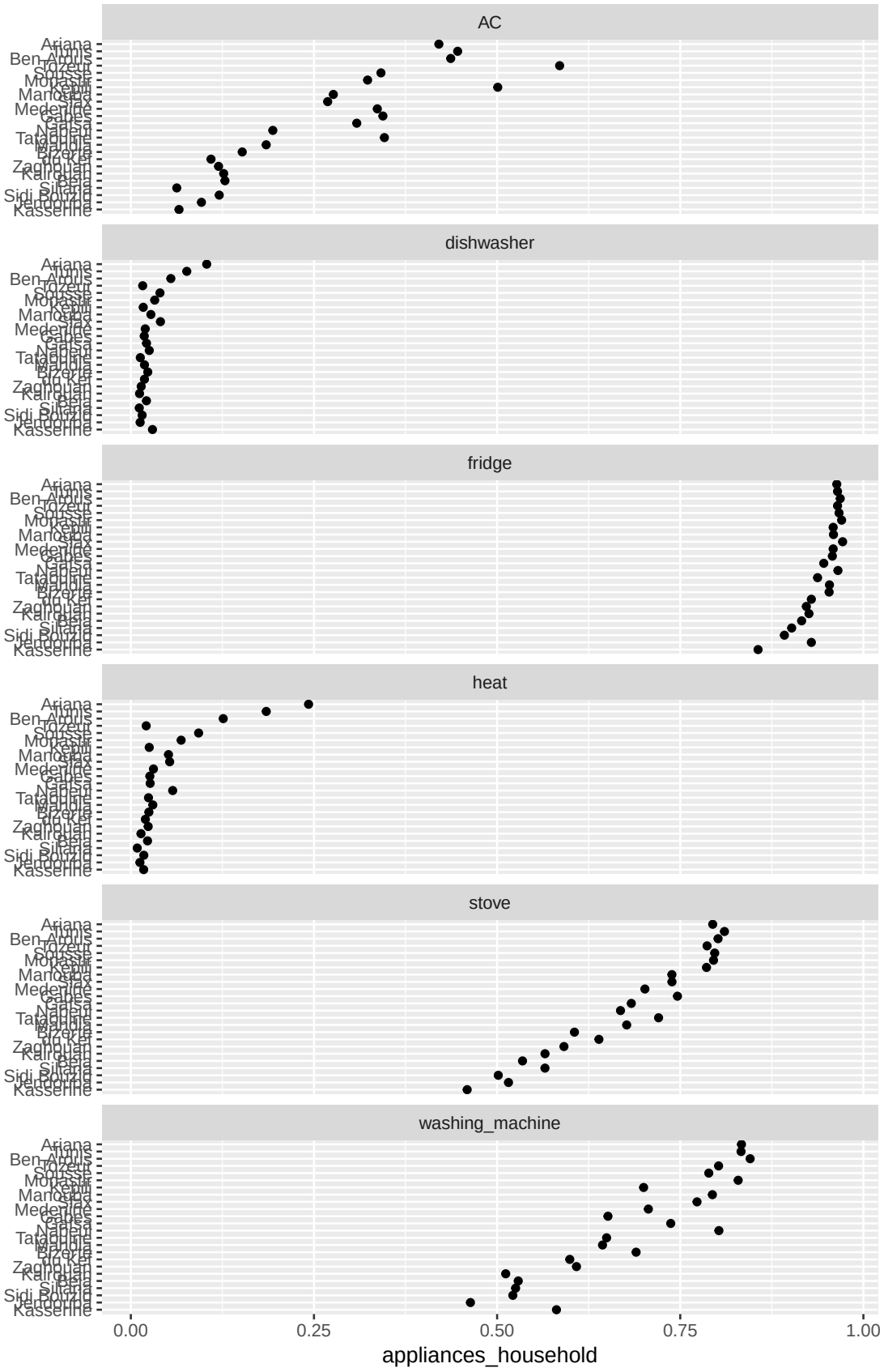
(6) Interactive component

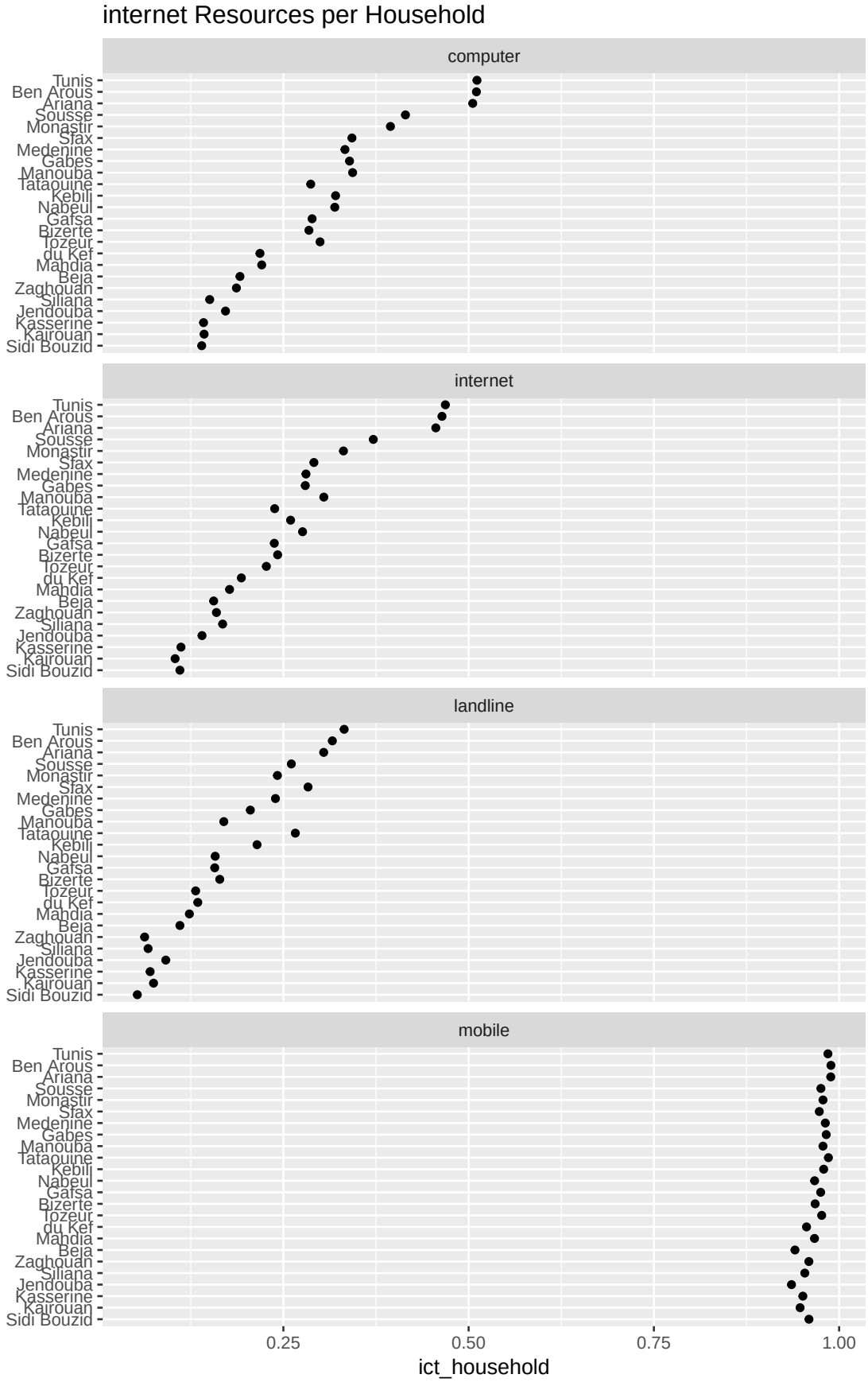
(7) Conclusion

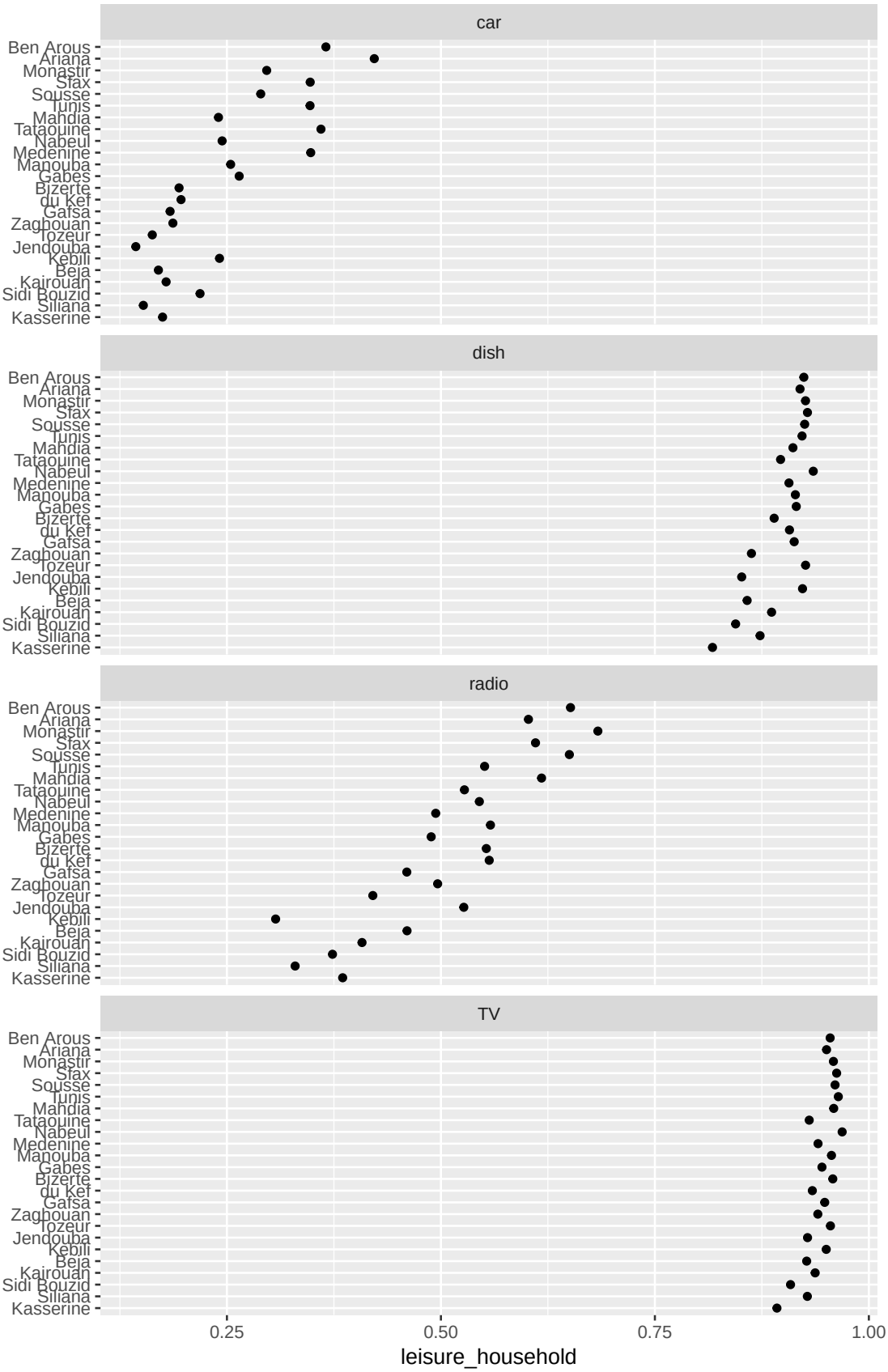
- Interested in looking at 2011 (year of the jasmine revolution) sicne the GDP went drastically down (Maybe less economic activity? Look into lumnisity?)
- Time series of luminosity per governorate
- !! Make a comment about coastal governorates
- https://academic.oup.com/eurpub/article/24/suppl_1/6/560448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

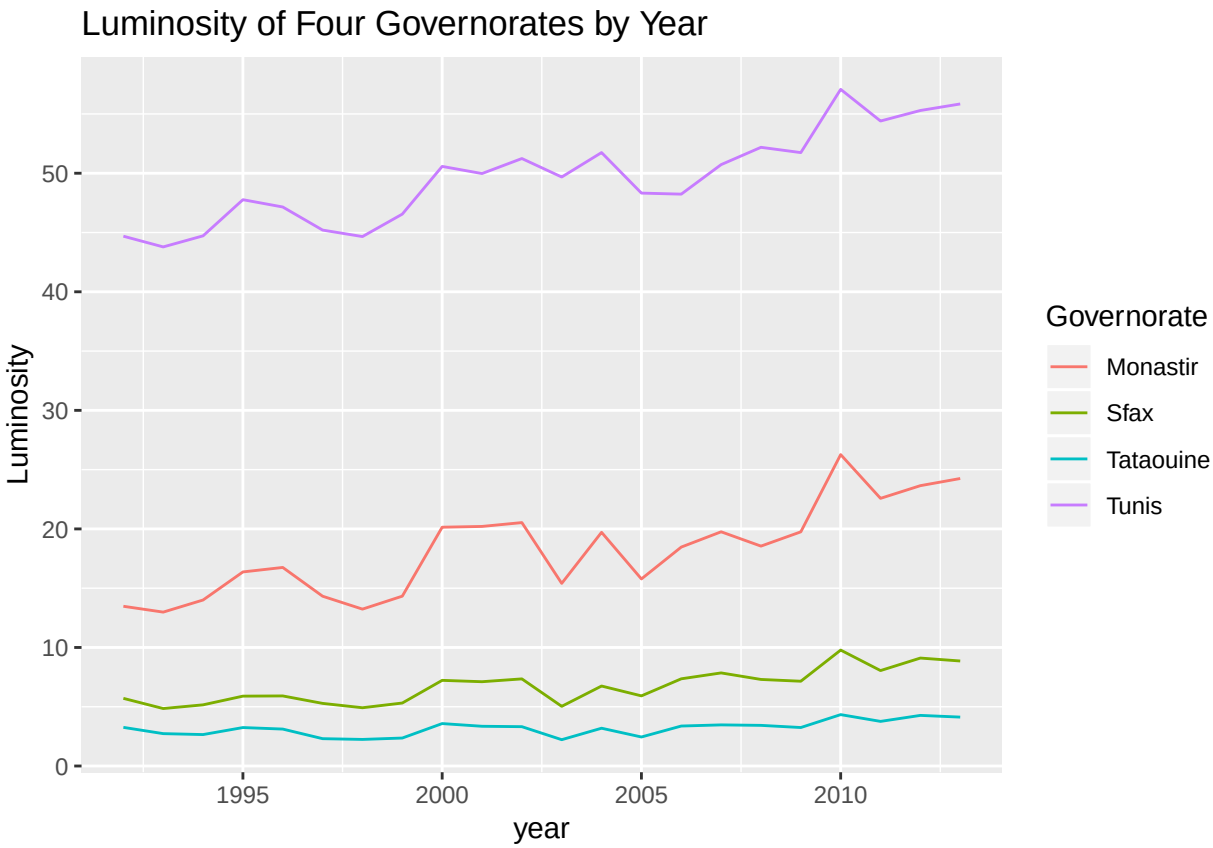
Appendix

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
AC
dishwasher
fridge
heat
stove
washing_machine
0.00
0.25
0.50
0.75
1.00
appliances_household


internet Resources per Household
computer
internet
landline
mobile
Tunis
Ben Arous
Ariana
Sousse
Monastir
Sfax
Medenine
Gabes
Manouba
Tataouine
Kebili
Nabeul
Gafsa
Bizerte
Tozeur
du Kef
Mahdia
Beja
Zaghouan
Siliana
Jendouba
Kasserine
Kairouan
Sidi Bouzid
0.25
0.50
0.75
1.00
ict_household


Leisure Resources per Household
car
dish
radio
TV
Ben Arous
Ariana
Monastir
Sfax
Sousse
Tunis
Mahdia
Tataouine
Nabeul
Medenine
Manouba
Gabes
Bizerte
du Kef
Gafsa
Zaghouan
Tozeur
Jendouba
Kebili
Beja
Kairouan
Sidi Bouzid
Siliana
Kasserine
0.25
0.50
0.75
1.00
leisure_household


Luminosity of Four Governorates by Year
Luminosity
0
10
20
30
40
50
1995
2000
2005
2010
year
Governorate
Monastir
Sfax
Tataouine
Tunis