2014

Exploratory Data Analysis

ASSIGNMENT III

AK

HCDE 511

Table of Contents

The Assignment

Assignment 3: Exploratory Data Analysis

Description

In this assignment you will use visualization software (Tableau highly recommended) to perform exploratory data analysis on a real-world dataset. The goal is to gain practice formulating and answering questions through visual analysis and to learn and critique a leading visualization tool.

You must work in pairs for this assignment. The assignment should be 3,000 – 5,000 words in length (10 – 20 pages with images). You'll be turning it in online by uploading it to the class dropbox.

Assignment

For this assignment, you will use a visualization tool to analyze a data set. We strongly recommend the use of Tableau Software, but you may choose an alternate tool (see below). If you do choose to use something other than Tableau, please contact us via email (hcde511@uw.edu) to make sure that it is appropriate for the scope of this assignment.

OVERVIEW

First steps:

Step 1. Choose domain & data

Step 2. Profile the data

Step 3. Pose questions

Iterate as needed.

Create visualizations:

Interact with the data

Refine your questions

For your write-up:

Keep a record of your analysis

Prepare at least one final graphic and caption to answer an interesting question

During your exploration of the data, we encourage you to create and record various types of views, including bar charts, scatter plots, maps and time series as appropriate for the data and question you are exploring. Note how different views support different questions and may reveal areas for further questions or exploration.

- Look at the data and/or its description. Write down at least four initial questions that you think the data may answer, including a comparative question, a correlation question, a geographically-oriented question, and a time-related or trend question.
- Use the visualization tool to examine the data for answers to your initial questions. You may wish to look for (for example)
  - relationships between pairs of variables (correlations, clusters)
  - outliers of various kinds
  - trends
- You may wish to refine your initial questions based on what you find as you explore. For example, you may wish to pose a related question about a subset of the data. Filtering, sorting, or other operations may be helpful.
- Use the visualization tool to explore the dataset and look for other unexpected kinds of relations. Note what features of the visualized data attracted your attention/focus. Try to highlight or otherwise isolate the subset of the data that contains an interesting feature.
- For some datasets it can be helpful to transform some of the data (e.g. by computing averages or medians, by converting numbers to percentages, etc.). You may do some of these kinds of transformations if you feel it is necessary or helpful (Tableau supports this), but if it is not needed than leave the data as is.
- Write up a discussion of what you found -- both expected and unexpected. This can include relationships that did not appear even though you thought they might. Try to report on at least one interesting or surprising piece of information. Be sure to illustrate your points with screenshots, but please scale them so they aren't too large.
- For each of your final questions, create a visualization that answers it. Be sure to label your axes and include an appropriate caption.
- In your discussion, comment on your use of Tableau (or other tool). What features did you find useful? Which ones were intuitive to use, and which were hard to understand? Was there any functionality that the tool did not have that you wished it did? In other words, how would you improve on the tool?

Data

Please use one of the following three datasets for the assignment. They should be sufficiently rich for you to be able to discover interesting information by exploring it. These sets contain a mix of nominal, ordinal, quantitative, geographical, and temporal data.

While these sets do not require a lot of pre-processing or coding on your part, the data may contain empty fields, spelling errors, or other incomplete or faulty data, which you may need to address as you explore.

We have intentionally removed some of the details about contents and size of each dataset below. Discovering the content and scope of your chosen dataset is part of your exploration process.

File formats: We offer each dataset represented as a CSV (comma separated values) text file. In addition to being readable by the visualization tools we will be using, it can also be directly imported into Microsoft Excel. You are encouraged to look at the data to get a feel for its contents, structure and scale before beginning your analysis. You may use Tableau to look at the underlying data after connecting to it (use the spreadsheet shaped button at the upper left under the word Data) or you may open the file in Excel or a text editor before loading it into the visualization tool.

Since each dataset includes a large amount of data, a large number of questions could be asked at many levels of detail. Using congressional candidate spending as an example, one might want to investigate spending and contributions at an aggregated level, breaking down the data by political parties at the national level. Alternatively, it is equally valid to filter out many of the attributes or entire sections of the data and explore, say, finances at a finer granularity, for example by investigating one's own state and local and neighboring congressional districts.

1. FAA wildlife strike data

This large dataset reports on collisions between wildlife and aircraft, based on pilot reporting collected by the United States Federal Aviation Administration. Each row of data describes a single collision between some type of wildlife and an aircraft. The data was downloaded from:http://www.faa.gov/airports/airport_safety/wildlife/database/. The form pilots use to report a wildlife strike may be helpful in understanding some aspects of the data: http://wildlife.faa.gov/strikenew.aspx

The data can be downloaded here in a compressed format that is optimized for Tableau:

- wildlife strikes v8.tde (17.7M)

or here for use in an alternate tool:

- wildlife strikes v8.csv (52.5M)

2. Congressional candidate spending

This dataset is a financial summary of U.S. campaign finance contributions and spending for each 2-year Congressional election cycle for a range of years. These data sets are published by the United States Federal Election Commission, and include summary information for every candidate for Congressional office (both the Senate and House of Representatives). The data was downloaded from: http://www.fec.gov/finance/disclosure/ftpsum.shtml

The dataset can be downloaded here:

- fec.csv (1.4M)

Each row of data represents a single candidate running for office in the given year (years are listed using the last year of the election cycle), and contains information about contributions, expenditures, party affiliation, state, congressional district (or none for the Senate), and outcome (win/loss/runoff). The dataset is fairly large. The only difference between the data set being given to you and the ones on the FEC website is that we have concatenated data for multiple election cycles and added a year column, enabling any number of trend analyses. As you'll undoubtedly notice, the data is highly multidimensional, with a large number of columns. The FEC website includes a detailed description of each column.

3. World economic data

This dataset is included as a sample dataset in Tableau 8.1. It includes a variety of measures about countries of the world over some recent years.

Here is a version of the data for use outside Tableau:

- world economic data.csv (0.5M)

Please provide a basic description of the contents, size (approximate # of rows and columns, length of time covered, geographical span included, rough number and/or range of dimensions) and quality (how complete, how many errors did you see, did quality issues impede your analysis) of the dataset you explored.

Tools

We encourage you to use Tableau for this assignment. However, you may also use other visualization tools of your choosing, such as: Ggobi, d3. If you use an alternate tool, please explain why in your writeup.

Grading (50 pts)

Assignments will be graded based on:

- Clear questions and applicable dataset
- Basic description of dataset contents, size & perceived quality
- The description of your visual exploration process
- Major view types included (bar charts, scatter plots, maps and time series) with appropriate related questions
- The depth of your analysis
- The design of your final visualizations
  - Instructive image (does it answer the question?)
  - Appropriate caption and description
  - Expressiveness/effectiveness of the visualization
- Comments and evaluation of the visualization tool including any improvements you might make.

The Solution

Analysis Log at a Glance

1. Start
2. Data Sets
    a. Went through the 3 different data set (cursory read)
    b. Chosen Data Set : **World economic data**
    c. Source : world economic data.csv (0.5M)
3. Opened in Excel (and Tableau) to get a feel of its contents, structure and scale
4. Initial observations
    a. Contents: World Economic Data
    b. Size:
        1. Columns: 59
        2. Rows: 2355
    c. Length of time covered: 10 Years (7/1/2000 to 7/1/2010)
    d. Geographical span included : Throughout the world (Africa, Asia, Europe, Middle East, Oceania, The Americas, Others)
    e. Number/ range of Dimensions: 4
        Country/ Region, Date, Region, Sub-region
    f. Quality:
        1. Completeness: 110/ 2354 Complete in terms of all the Measures. I.e. all of these had values for all the measures that had some value anywhere in its column.
        2. Foreseeable issues: Complete data size (for Health) is small, which may induce inaccuracies in interpretation. Also, the geographic region with maximum number of records is Africa, which is at 594/ 2354.
5. Tentatively chose Africa (594/2354) as the geographic region under consideration.
6. Focusing primarily on: Health data [1967/2354 (total, 100% data completeness in Health measure)], pertaining to Arica (561/ 2354), and its relationship with its attributes and dimensions, and that of others.

Core Analysis Steps:

7. During all this, profiled the data, and found 5 initial questions to pose.
8. Decided to focus on Africa. Created visualizations to understand what's going on a larger scale.
9. Created targeted visualizations and took snap shots.
10. Probed and refined questions further.
11. Iterated a few times (Steps 7 to 10).
12. Prepared final graphics to answer the final questions.
13. Performed analysis as required. Included them in appendix as appropriate.
14. Logged shortcomings of the visualization software (Tableau) as and when encountered.
15. Cited/ referenced as appropriate.
16. End.

Exploratory Data Analysis

Exploratory Data Analysis is one of the two main types of analysis (the other being the Directed Analysis) in which the analysis begins without a set of targeted questions. In this analysis, one first tries to get his/her bearings in the dataset, and then based on this and his/her area of interest, he/ she decides to iteratively proceed ahead in a fashion similar to Directed Analysis until a set of pertinent and critical questions (and hopefully the answers to them) is obtained (Few, 2009).

This, of course, is theory.

For this assignment, I had 3 data sets to choose from, out of which I choose the **World Economic Data**, which included a variety of measures or parameters about the different countries of the world spread over a decade. I started with order and structure of Exploratory Data Analysis, but it soon turned into almost a classic example of the Chaos Theory before aligning into Directed Analysis.

Iteration 1

Initial Observation, Analysis and Description

After choosing the data set, it was opened initially in MS Excel to get a feel of its contents, structure and scale. This data set had several measures and dimensions, and contained a mixture of nominal, ordinal, geographical and temporal data.

In terms of contents, this data set consisted of data points pertaining to the world economic data under different types of measures, which were classified under Business, Finance, Health, Military & Government, Population, Stock Market, and Travel.

Overall, the data set spanned across 59 columns and 2355 rows, covering 10 years of time from 7/1/2000 to 7/1/2010. In geographical sense, the data was divided as per regions (continents) almost throughout the world, and sub regions (major parts inside a continent), which could be drilled down to individual countries. The number of dimensions used to encompass all this was 4 (Country/ Region, Date, Region, Sub-region), thus providing with a wide spread.

However, in term of quality, the data set seemed to be lacking a bit, in both completeness and volume. During initial observations, it was found that only 110 of 2354 records were complete in terms of all the Measures having a value for each of the records, which reduced the data set under consideration substantially if analysis of complete and correct records was desired.

The issue in this case could have been that though all of the points on the graph would be accurate in their portrayal of the individual units of information that they represent, all of these together could be of limited value in terms of finding significant trends and correlations between variables, or in short, gaining an effective big picture. Further, this could induce inaccuracies in interpretation.

While sifting through data, it was observed that the data set for the Health-measure was the most voluminous with respect to choosing data completeness (or minimal number of errors) as a priority/ requirement. Further, after assimilating the global picture, the scope was narrowed down to a region where more pertinent questions were posed. I reasoned that as I was dealing primarily with Health, the parameters that constitute it would effect it differently in different geographical areas, and hence a generalized solution might not be useful. The following link to the video also helped me to understand and rationalize it better.

Link:http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen?awesm=on.ted.com_toUI&utm_campaign=&utm_source=facebook.com&utm_content=awesmpublisher&utm_medium=on.ted.com-facebook-share/

Following this, I found the geographical region with the maximum number of records (594/ 2354) to be Africa, from where I chose to commence my more focused analysis (**Iteration 2**). This was because of the highest volume of quality data available, strong "presence" of health related measures here (Africa, as depicted below), and my interest in the Health sector in this region.

Initial Questions

After this initial data profiling, following are the initial questions that I decided to pursue:

1. In which countries has the infant mortality worsened with the progression of time?
2. Which regions in the world are the unhealthiest in terms of infant mortality?
3. Are there any outliers in terms of the change in life expectancies in the world during the years 2000 to 2010?
4. Are life expectancies and infant mortality correlated?
5. How do the female and the male life expectancies depend on each other?

Visualizations and In-Depth Analysis

Some initial visualizations that helped me gain a clearer picture and make the aforementioned decision about focusing on the African continent for this assignment are as follows (these are mostly ordered according to my initial questions):

In which countries has the infant mortality worsened with the progression of time?

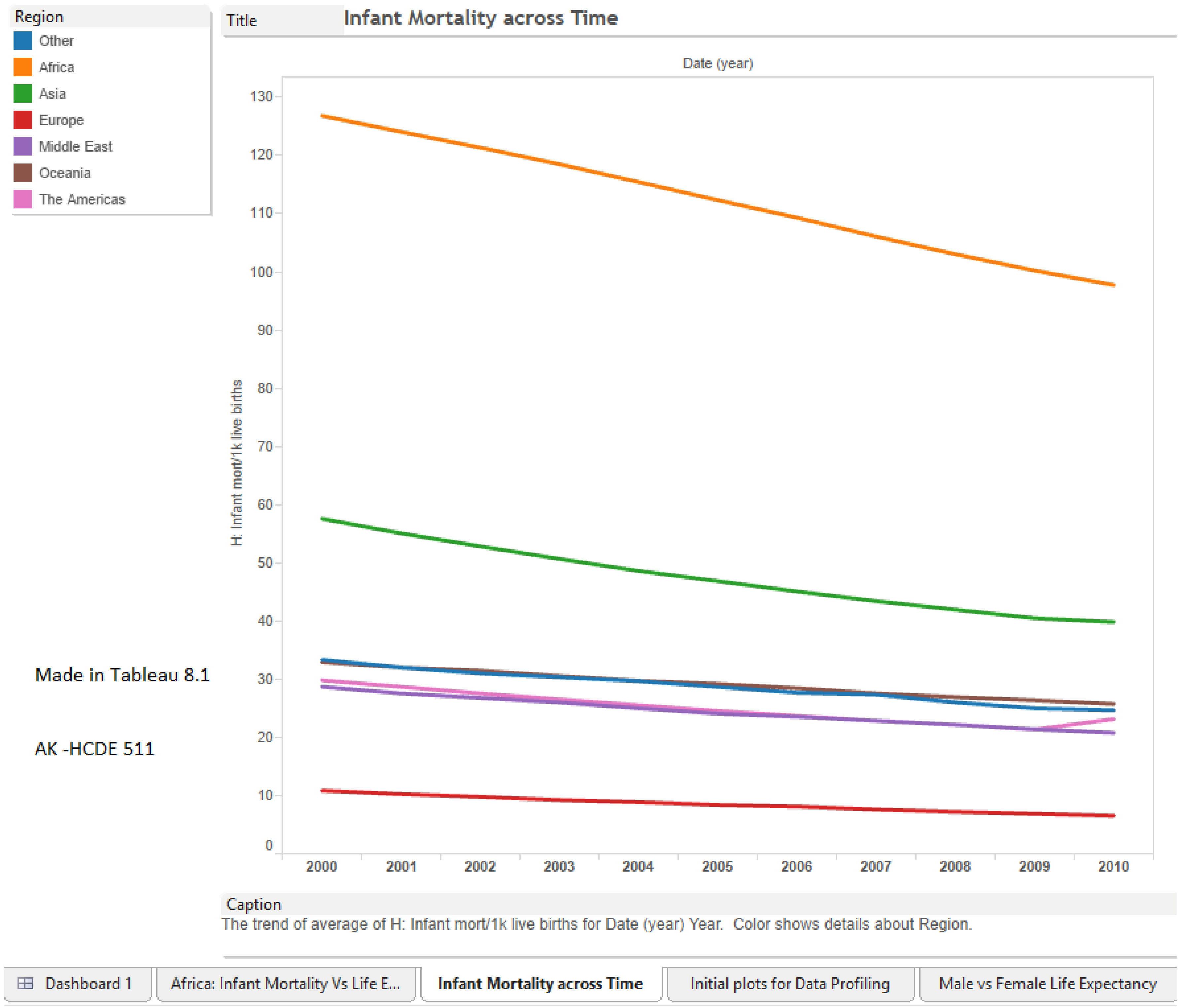

Fig. 1

In this graph, the steepest reduction (in terms of numbers) in infant mortality seems to be in Africa. This also depicts the severity of infant mortality in Africa as compared to other nations (almost twice as severe as the next region (Asia)).

This graph also answers the question regarding the regions where the infant mortality has worsened– The Americas (which is also an outlier here), which is surprising as this region contains the United States as well. To explore this further, I made the following:

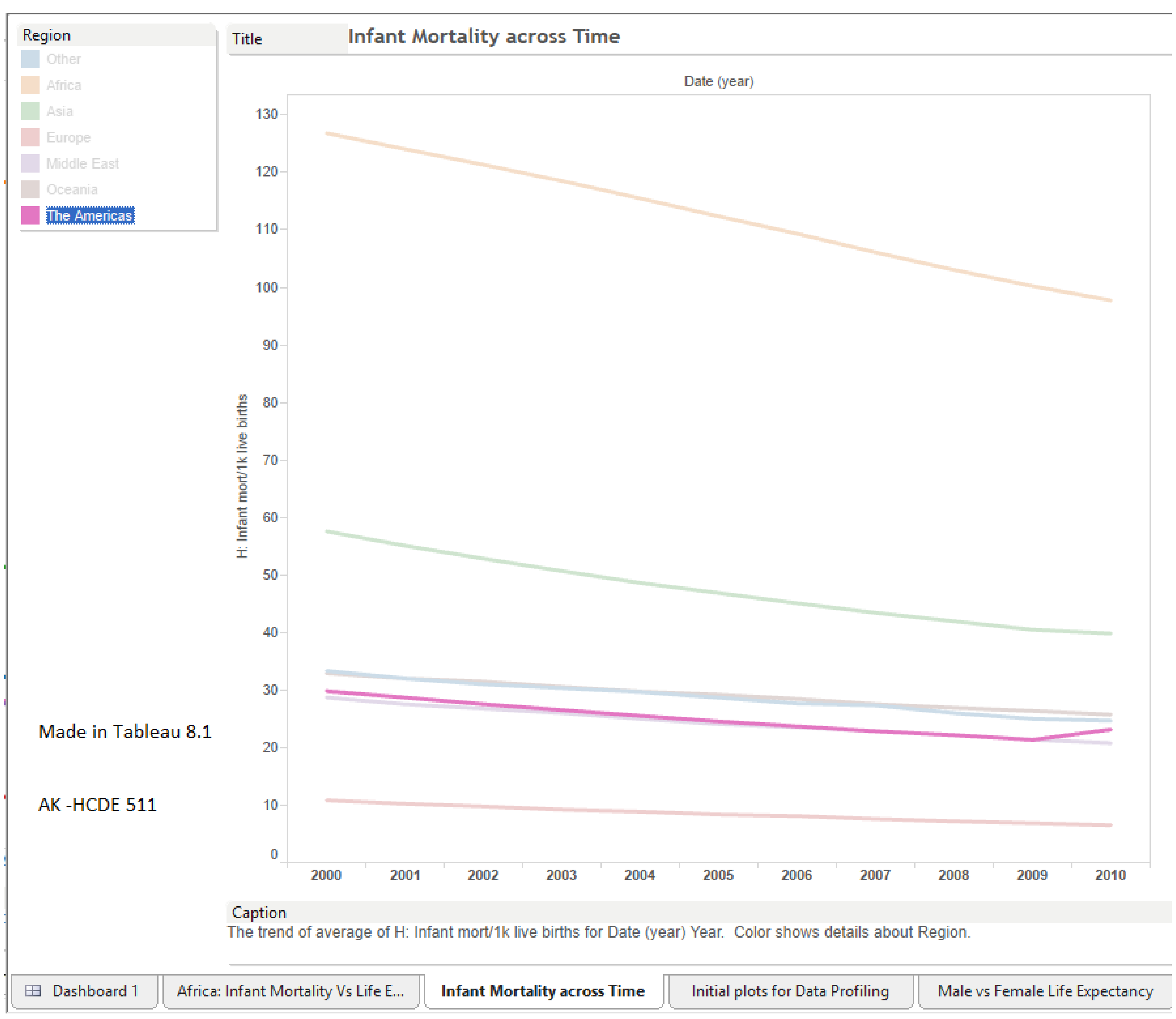

Fig. 2

This graph shows clearly the rise in infant mortality in the Americas during 2009-2010 when compared to its recent past. Upon drilling down a bit, we get:

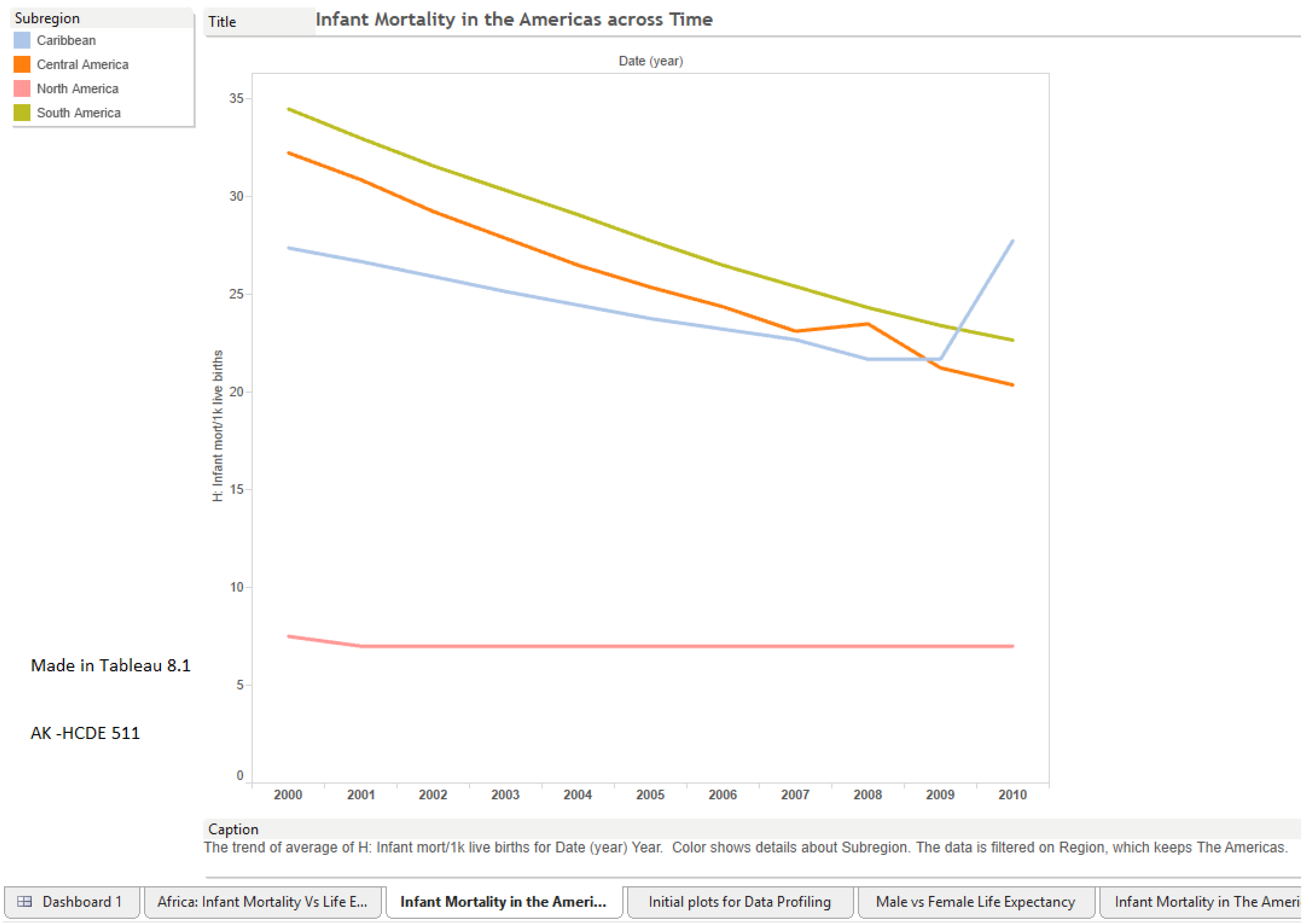

Fig. 3

This is interesting as it allows us to examine closely the sudden rise in the infant mortality in Caribbean, which is in stark contrast with all other regions on this and the previous graphs, were the infant mortality has been steadily decreasing, or has been maintained at a low value (North America).

And as for North America, though the infant mortality rate here is lower than other regions, there has also been no change in it. So perhaps, one could conclude that either it has reached the point of natural saturation where the mortality rate cannot go down any further with the current technologies and knowledge, or it could be because of something else which is causing all the efforts in this arena to become ineffective, or go towards channels not intended. It could also ascribe to the rise or occurrence of new things like disease, pollution, etc. which might be affecting this, and maintaining the infant mortality rate instead of letting it reduce.

Another observation here is that before the steep rise in the infant mortality in the Caribbean, there was also a rise in infant mortality in Central America, which is well connected with the Caribbean. This could indicate that there could be a connection (a possible causation) between the two (e.g., a disease/ epidemic that started in Central America, and got communicated to the Caribbean). Though this may need to be substantiated well – as the North and the South America, which are also well connected with the Central America, don't seem to be affected.

Which regions in the world are the unhealthiest in terms of infant mortality?

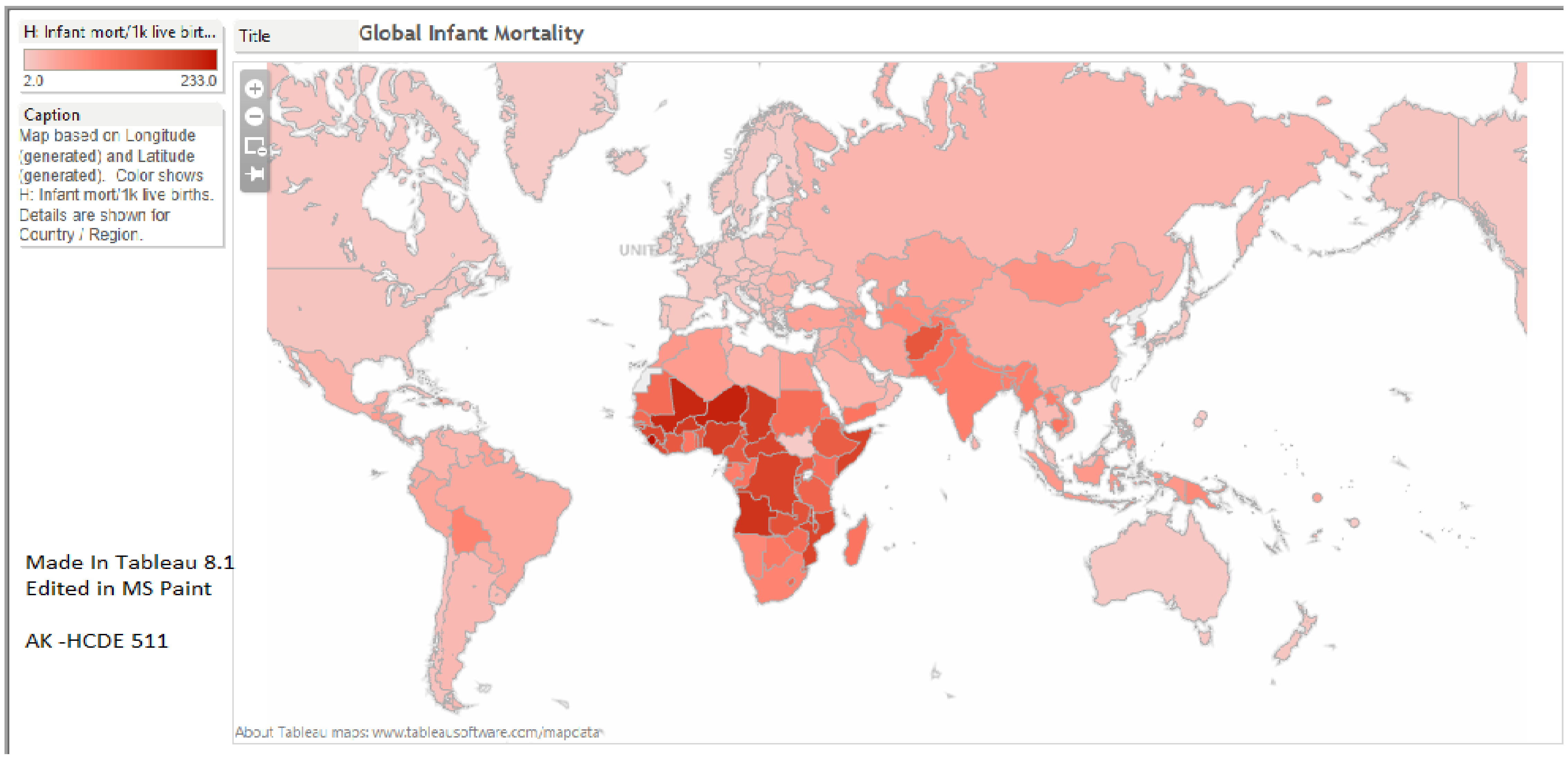

Fig. 4. This graph shows the severity of infant mortality across the globe, with the more severe ones depicted with darker reds.

To understand the distribution of this severity on a deeper level, I employed a filter with the infant mortality rate above 100. The following depicts the visualization thus created:

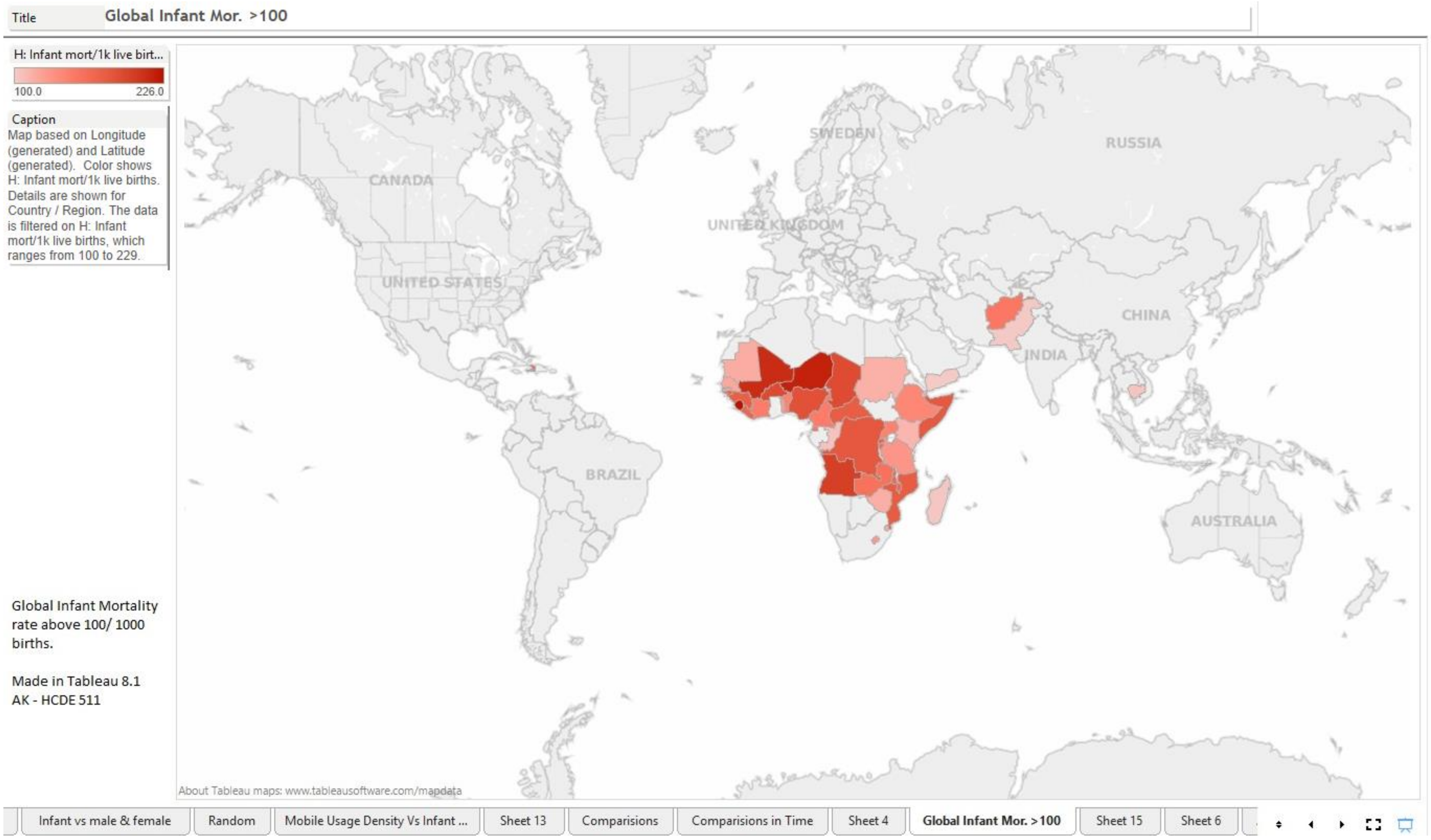

Fig. 5

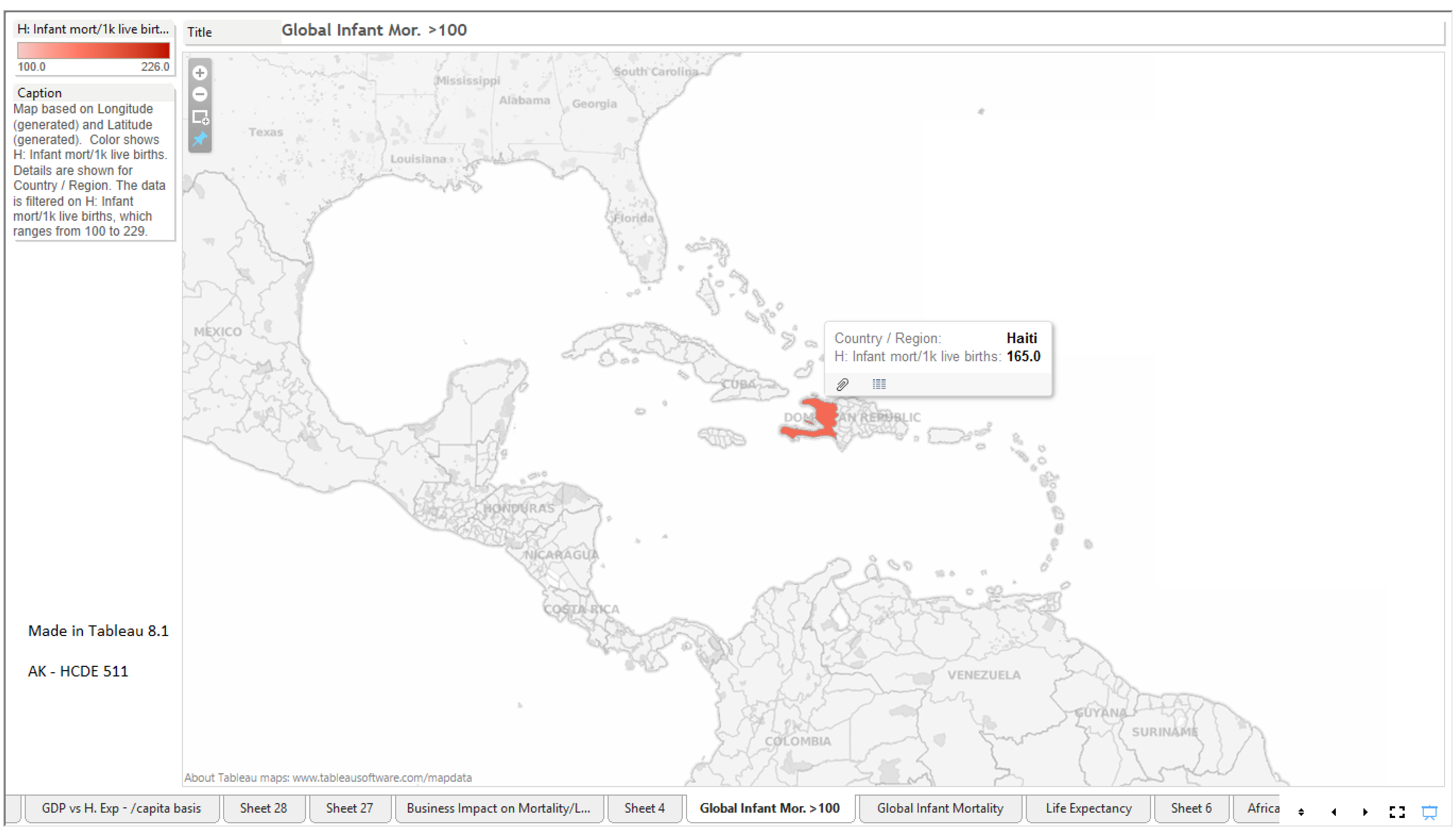

Fig. 6

As can be seen in the above 2 graphs, the infant mortality is most prevalent in the African continent, both in terms of severity (rate) and size.

Thus, along with the comparatively voluminous data available for Africa, it's profiling done so far (via Excel and Tableau), and the nature of the above visualizations, I selected Africa as the prime scope of consideration for this Assignment. I believe that the data concerning this would help in painting a vivid picture regarding its underlying intricacies and would aid in perceiving patterns and trends faster.

Are there any outliers in terms of the change in life expectancies in the world during the years 2000 to 2010?

The following visualizations address this effectively:

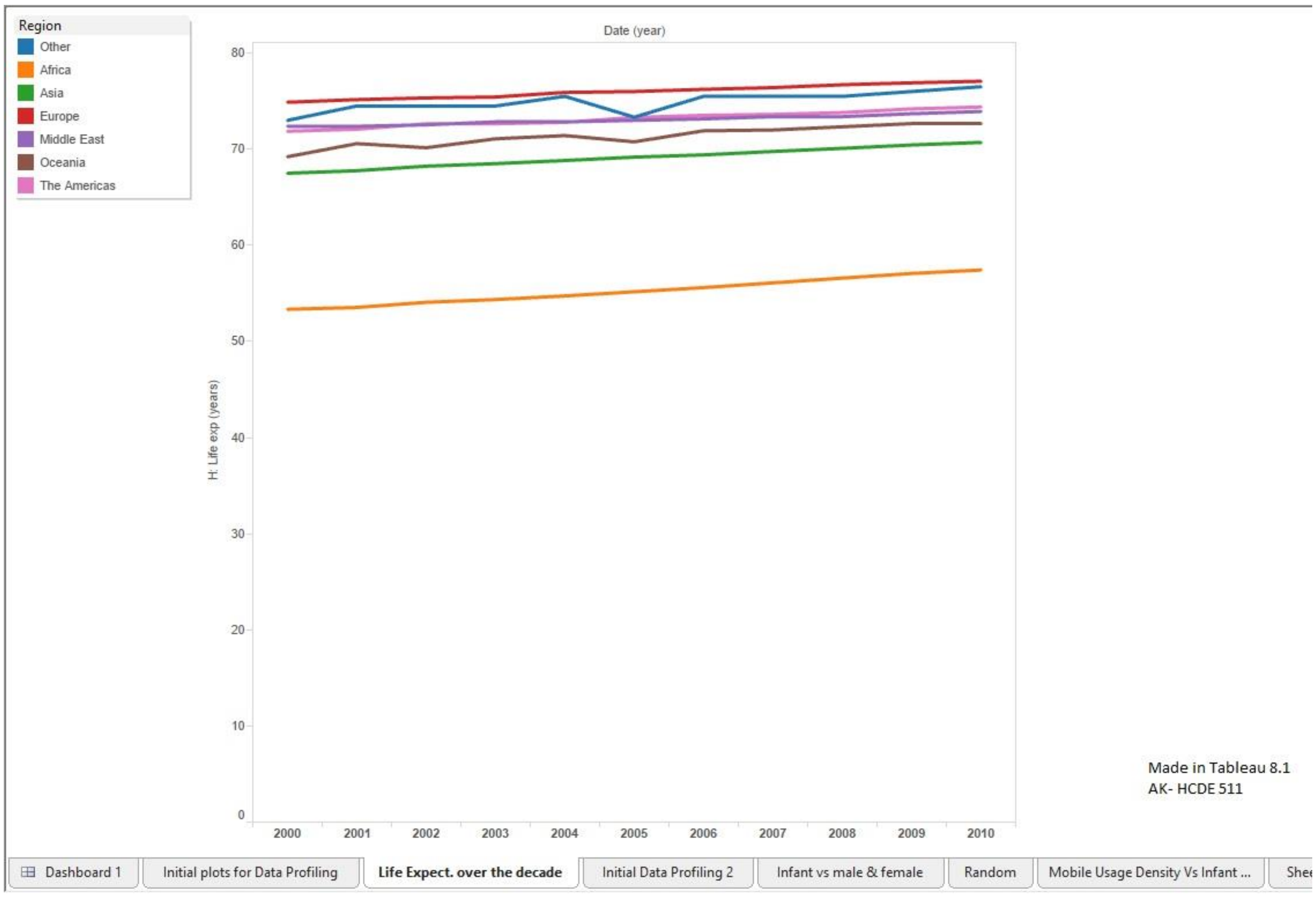

Fig. 7

Here, across the decade of time, all the regions have been moving towards higher life expectancy.

The African continent could be called an outlier among them in terms of the numeric value of its life expectancy. Though the life expectancy seems to be increasing at the same rate here when compared to the other regions, the value of its life expectancy itself is much lower as compared to others (at least 10 years less than the next best region in terms of life expectancy).

To get a clearer picture, I drilled down to find:

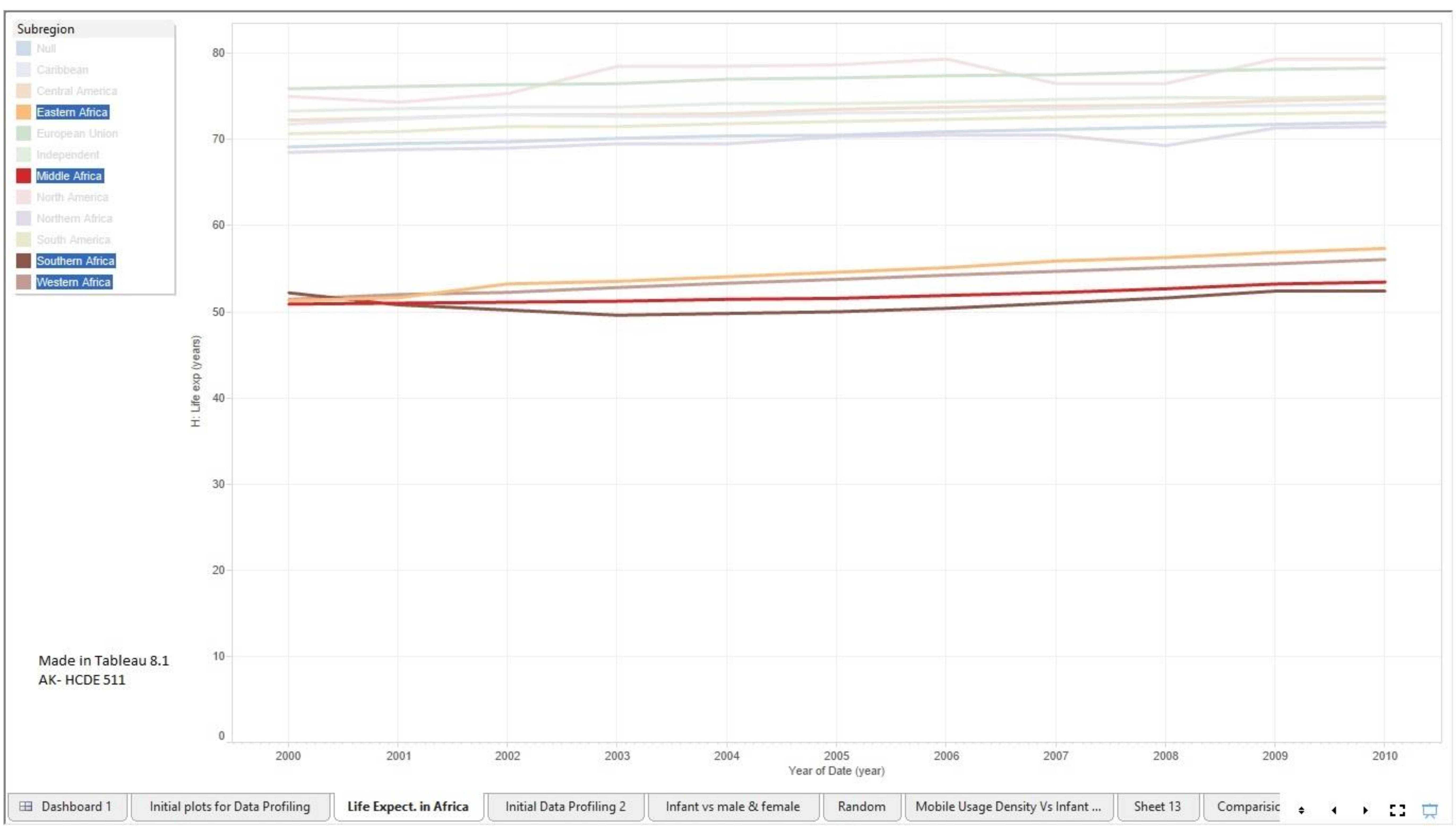

Fig. 8. Aggregate measures with respect to the different sub-regions focusing on the African continent.

For additional insight, I created the following which displays the aggregate measures of life expectancies of countries, color coded by Region.

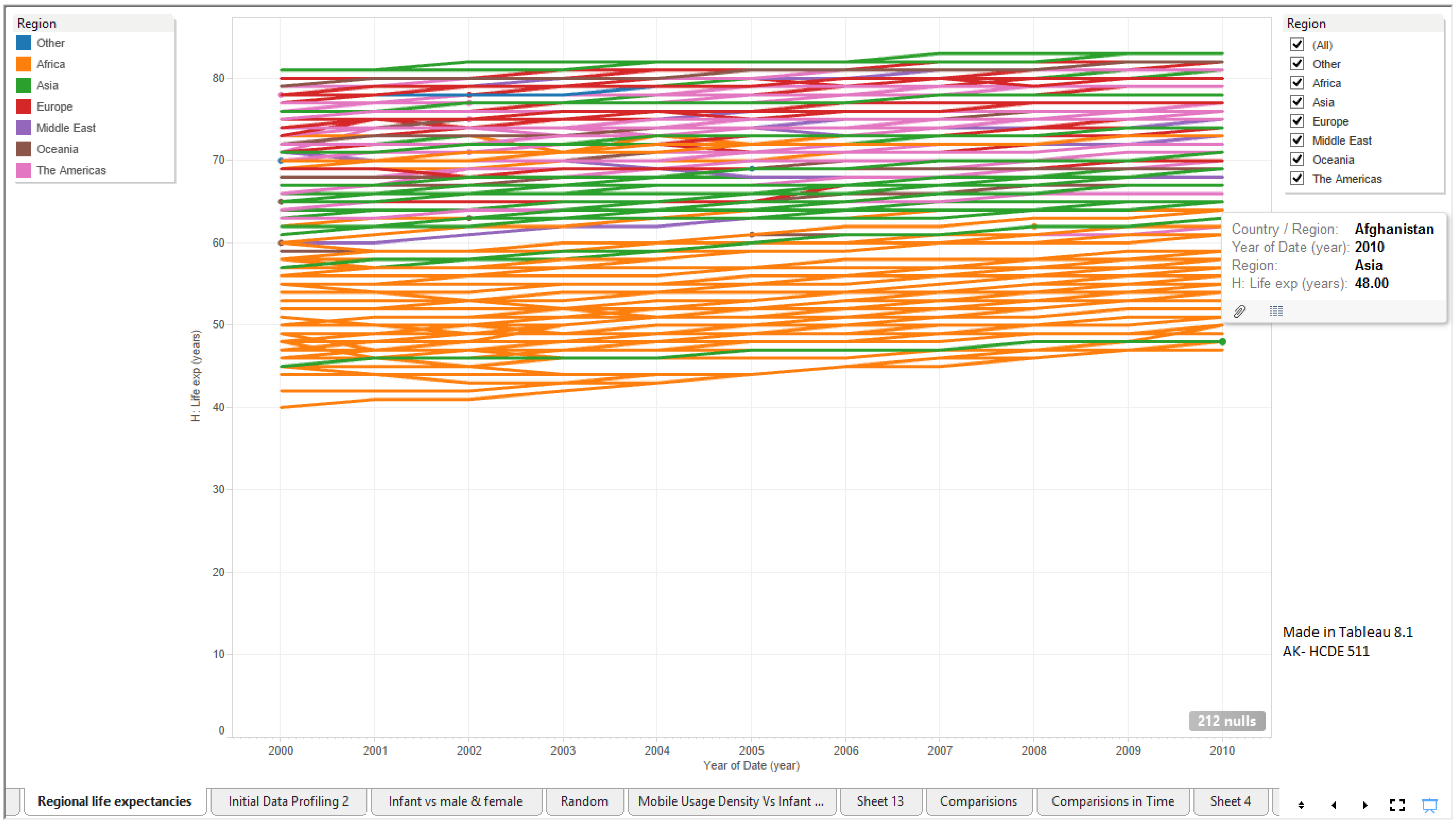

Fig. 9 Aggregate measures with respect to the different countries (globally), color coded by region.

As is further elucidated here, the majority of the African life expectancy falls between 40 years and 65 years, which is below almost any other region.

Also, another outlier becomes apparent here – Afghanistan, which is represented by the green line near the 45 year life expectancy mark.

Focusing on the data pertaining to Africa, and segmenting it into sub-regions of Africa, we get the following:

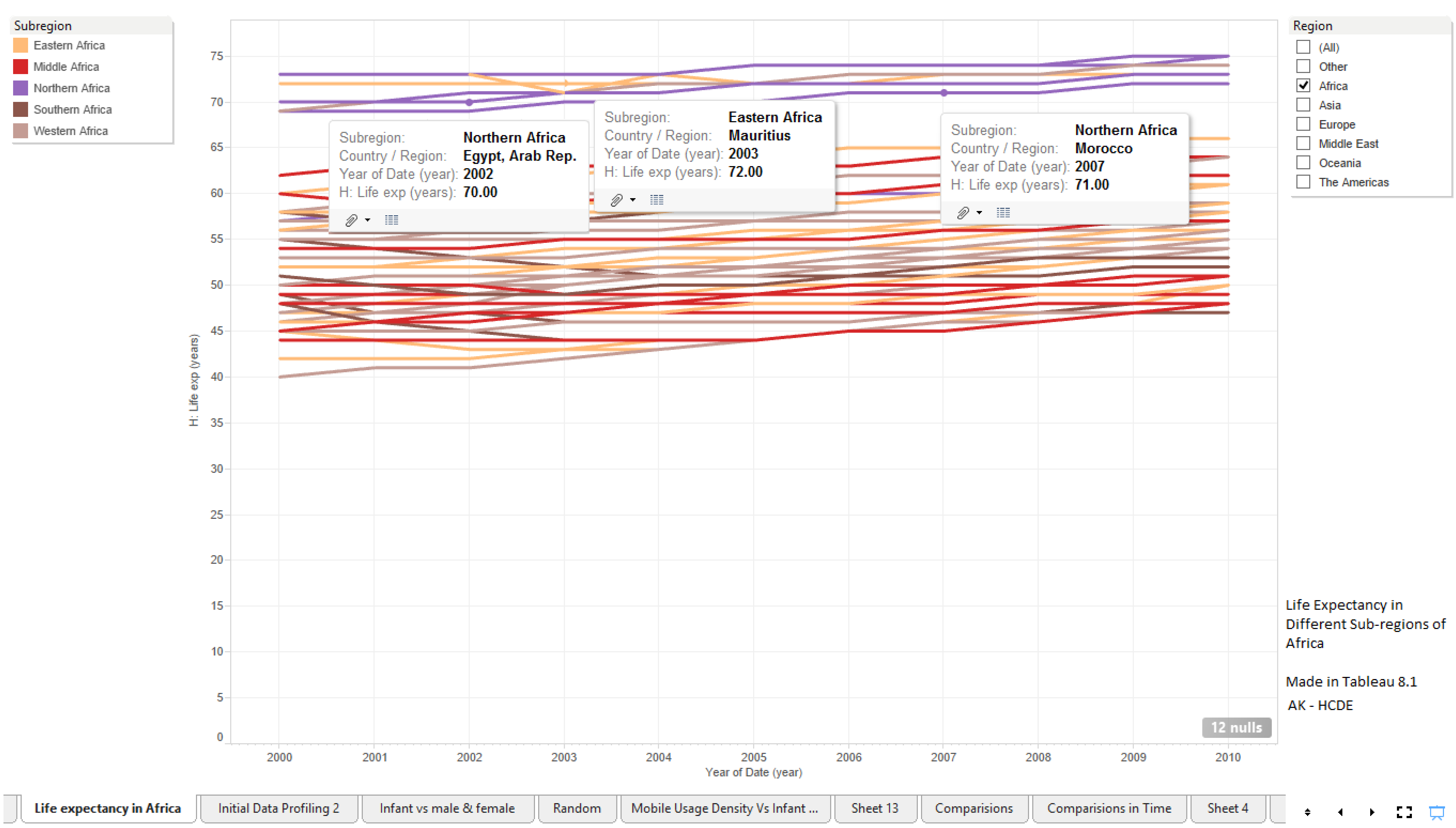

Fig. 10

As mentioned before, the life expectancy in the African continent is towards the lesser side compared to the global spectrum. However, this graph shows that there are some parts of Northern Africa (and Eastern Africa, to a limited degree) which have a much higher life expectancy (as depicted by the data points in the graph). Upon closer inspection, these places are typical touristy and "modernized" locations, with higher amount of wealth.

Hence, it would be interesting to investigate the possibility of a correlation between life expectancy (and possibly infant mortality) with respect to the tourism industry, and the usage of internet and mobiles (which are few things in the dataset that are known to be good estimators of "modernization" of a place). Further, another important topic worth investigating would be the relationship between the money spent in the health sector, versus its impact on infant mortality and life expectancy, which we shall revisit in **Iteration 2**.

On a side note, as we were dealing with graphs concerning life expectancy and infant mortality, I was curious to see how the Caribbean is faring in terms of life expectancy in the later spectrum of years as it had a sharp increase in its infant mortality rate then. The following graph addresses this:

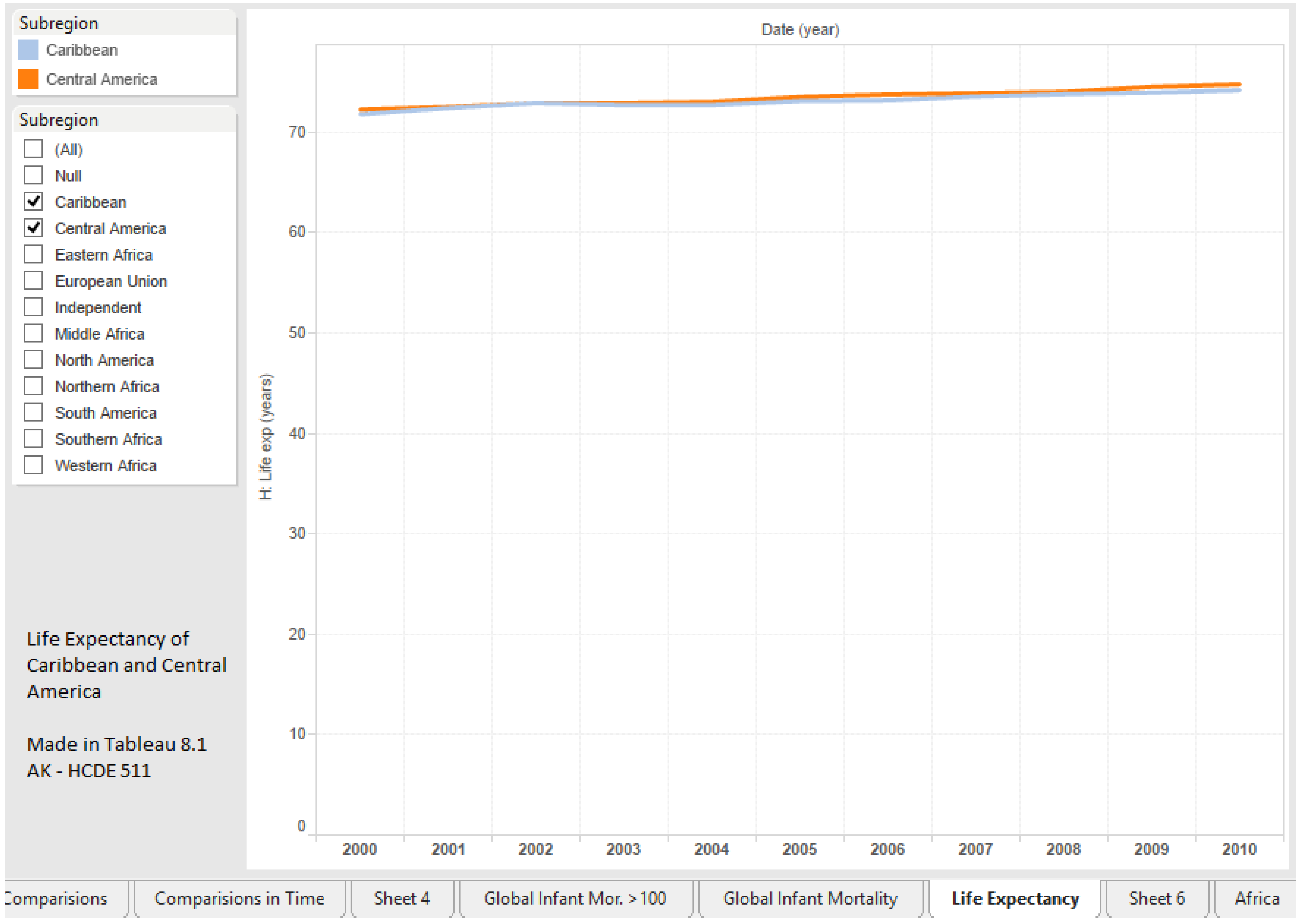

Fig. 11

As can be seen here, this graph doesn't indicate any dip in the life expectancy of either of the two regions. According to Google, life expectancy is the *average* number of years a person will live. This means that if there was an increase in the infant mortality, there would also be an increase in the number of years the surviving people were living to keep the life expectancy near constant. In simpler terms, if the number of deaths pertaining to younger age group increases, then the approximate age of death for other people (which need not be on the comparatively older age of the spectrum) would have to increase, so that the life expectancy of that region is being maintained or increasing, as is happening in these areas as shown in the above graph.

Here's a paragraph from Wikipedia:

"It is important to note that life expectancy is an average value. In many cultures, particularly before modern medicine was widely available, the combination of high infant mortality and deaths in young adulthood from accidents, epidemics, plagues, wars, and childbirth, significantly lowers the overall life expectancy. But for someone who survived past these early hazards, living into their sixties or seventies would not be uncommon. For example, a society with a life expectancy of 40 may have very few people dying at age 40: most will die before 30 years of age or after 55."

Source: Life expectancy - https://en.wikipedia.org

So, an interesting thing that could be addressed here would be finding the reason that could be causing this apparent continued improvement of life expectancy.

Are life expectancies and infant mortality correlated?

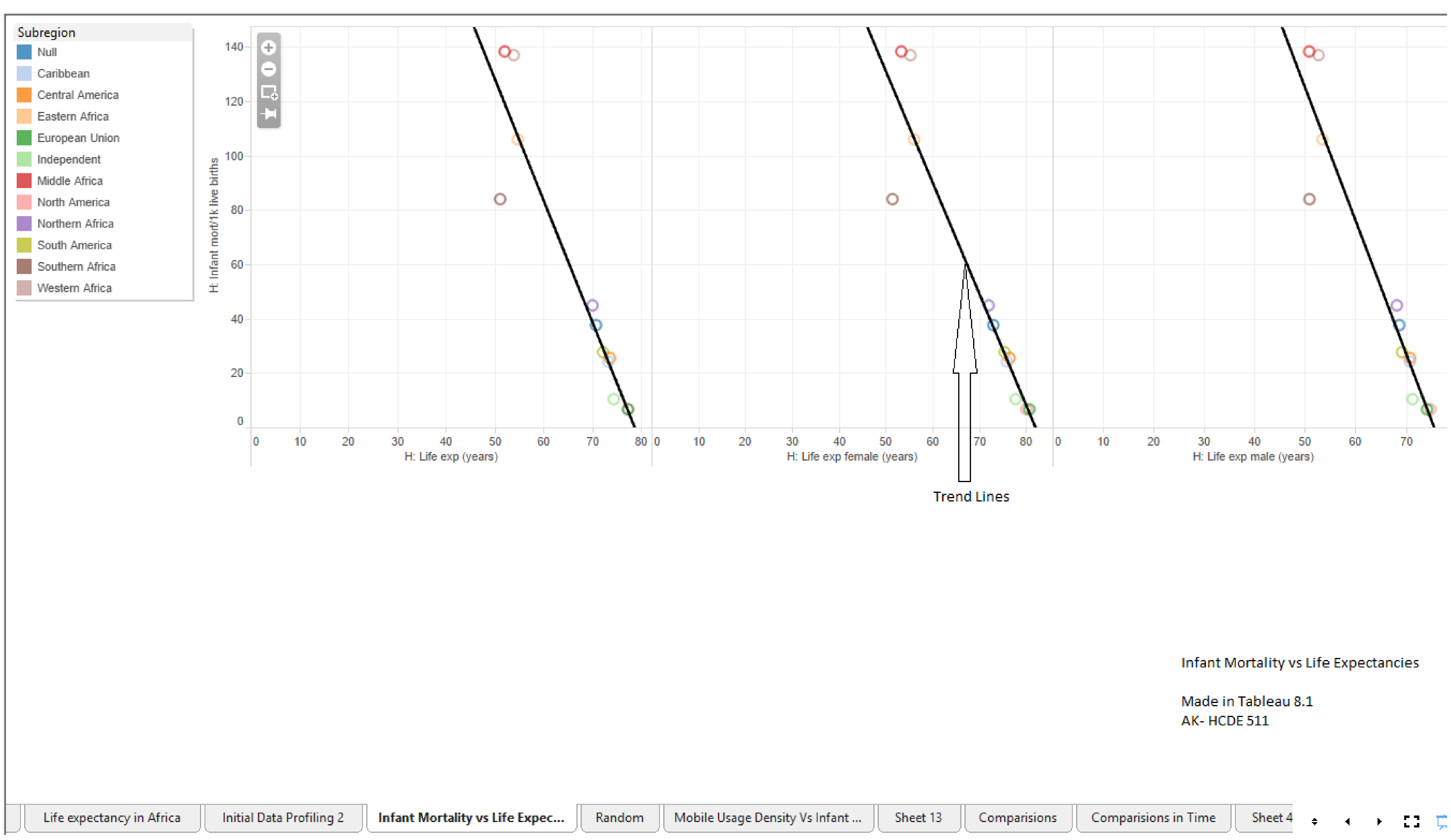

Fig. 12

On a global scale, the infant mortality and life expectancy (all, female, male) seem to be inversely correlated, as is depicted by the scatter plot and the trend lines in it. Further, the infant mortality seems to be more sensitive to male life expectancy (as per the slope of the graph).

For more clarity, here's the non- aggregated data-points version:

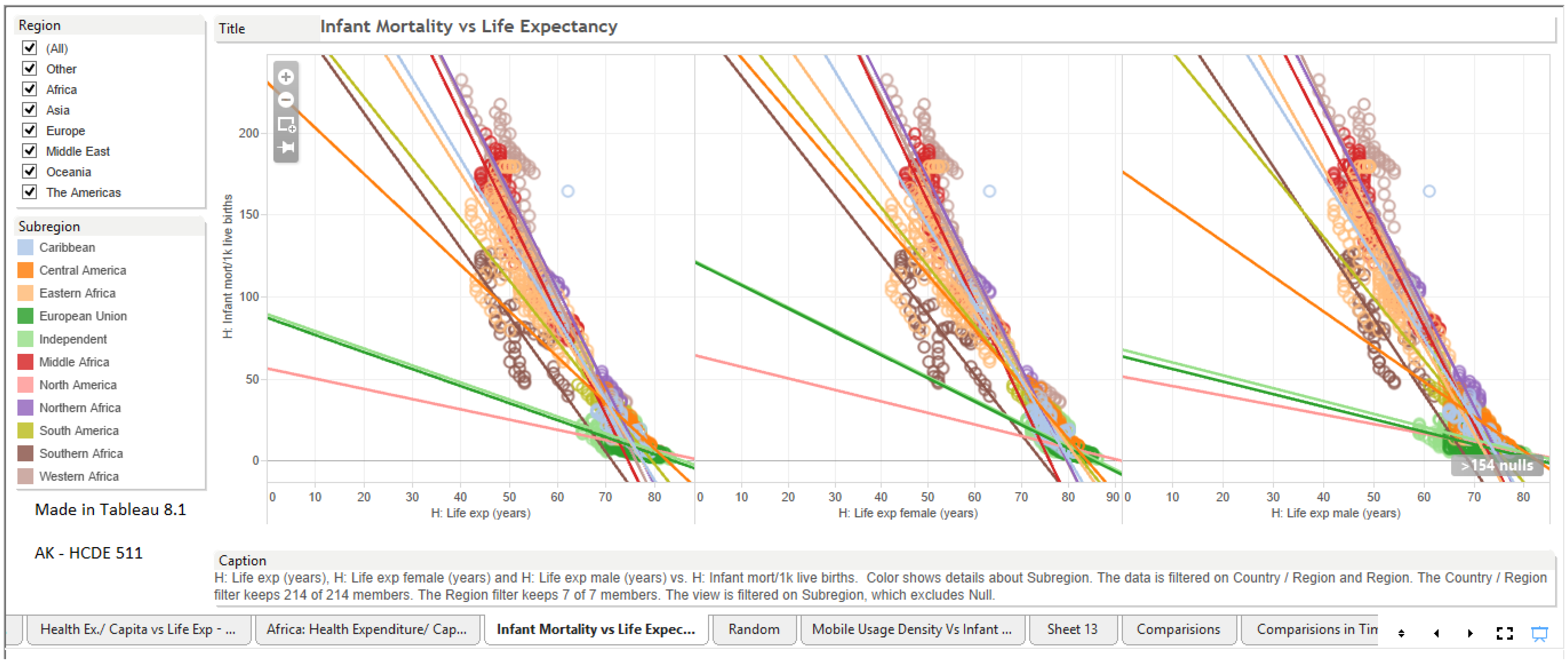

Fig. 13

As can be seen in the above graph, the trend lines seem to be very distinct, most of them going through the highest data-density area in the graph.

Further questions that could be worth investigating based on this and our area of focus would be in understanding these measures across the time interval of the 10 years, and the associated trends, if any, especially pertaining to the African continent (**<u>Iteration 2</u>**).

How do the female and the male life expectancies depend on each other?

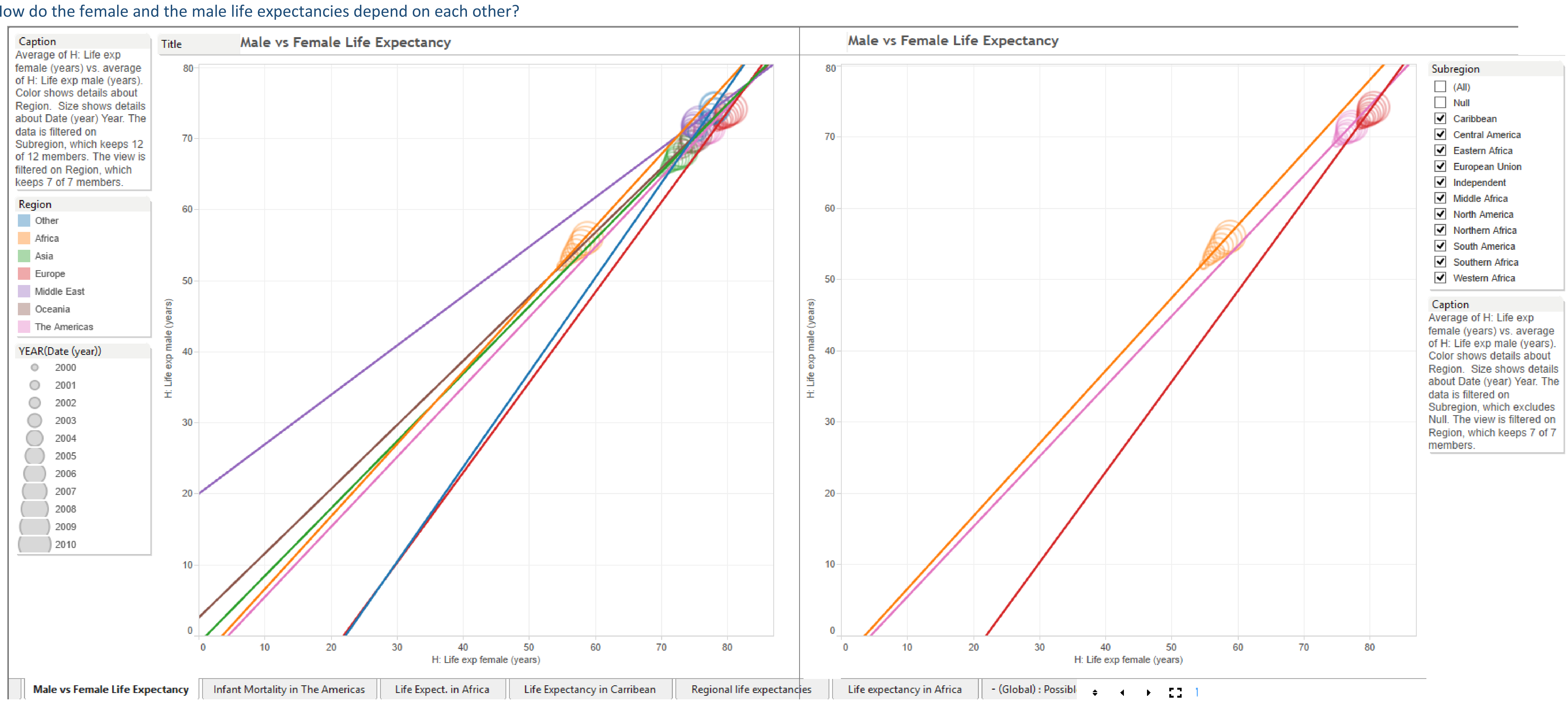

Fig. 14(a)

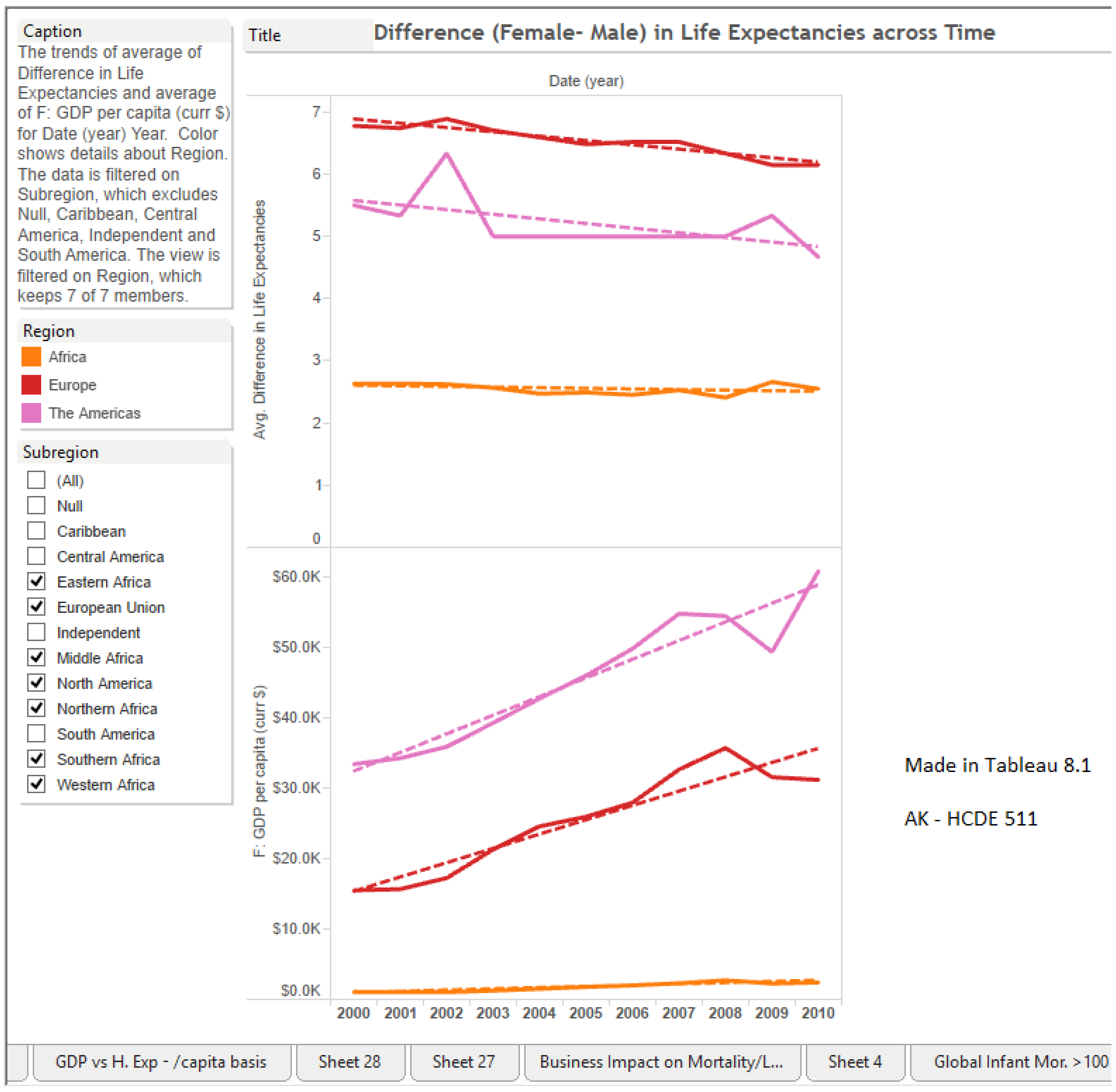

Fig. 14 (b)

Interesting Observations here:

1. Male and female life expectancies seem to be directly related, with extremely strong trends.
2. Females tend to live longer than males. As the countries get richer, the gap between their life expectancies tends to reduce, though both male and female life expectancies continue to increase.

This could imply that females are more resilient in dealing with poorer/ adverse/ more difficult living situations than their male counterparts. And as conditions improve, males seem to get closer to female life expectancies, while the latter continues to increase till they reach a saturation point.

Another aspect could be that because females live longer, they are closer to reaching their optimum saturation point of life expectancy, compared to males, whose life expectancy needs to keep increasing for a while before they reach their optimum saturation point. So, when we move towards that point, it might appear that the better living conditions seems to be effecting males better than the females. This can be rectified by pulling up the actual values of life expectancies on the graph at that stage.

Moving on, the latter concept would form a question for future iterations, especially in the African continent's context, wherein we would be investigating the same parameters for the economically richer parts of Africa with respect to the comparatively poorer ones. (**Iteration 2**)

Aside from this, I made a few visualizations to get a feel of the dependencies of Infant Mortality:

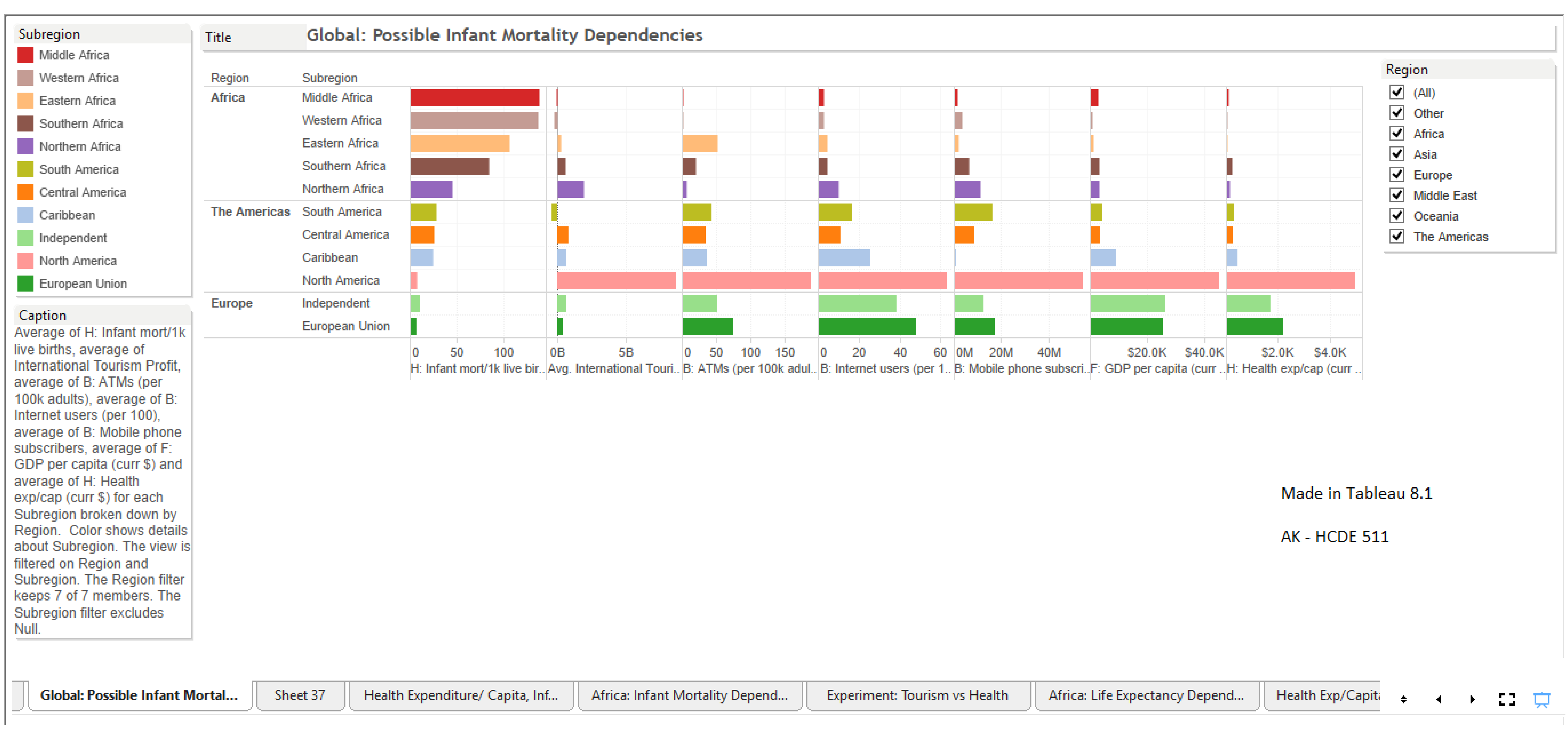

Fig. 15

In the above graphs, it seems that the infant mortality has a strong inverse relation with most of the technology and financial measures. Upon digging further, we find the following:

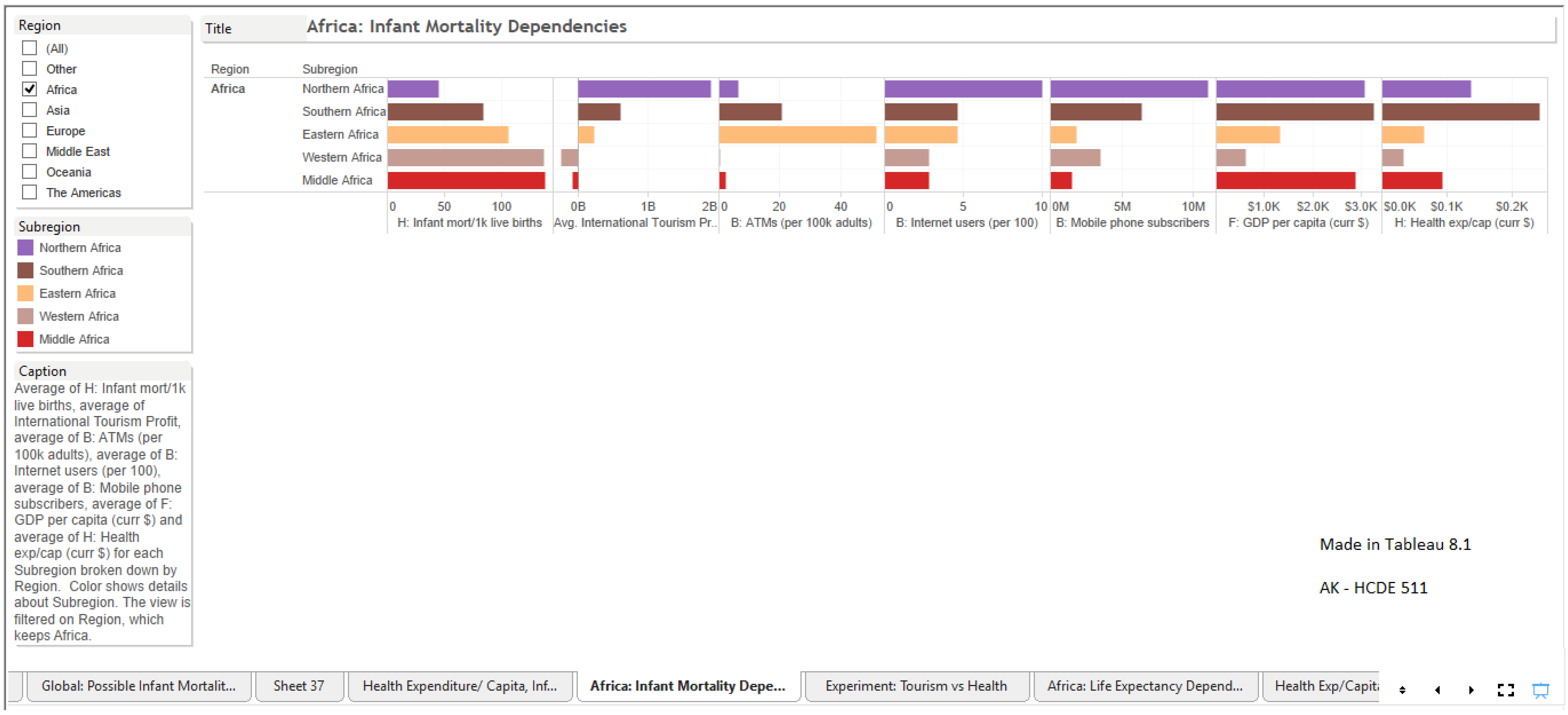

Fig. 16

It appears here that the Profits earned from international tourism are also inversely correlated with infant mortality.

The following scatterplots provide more clarity in this:

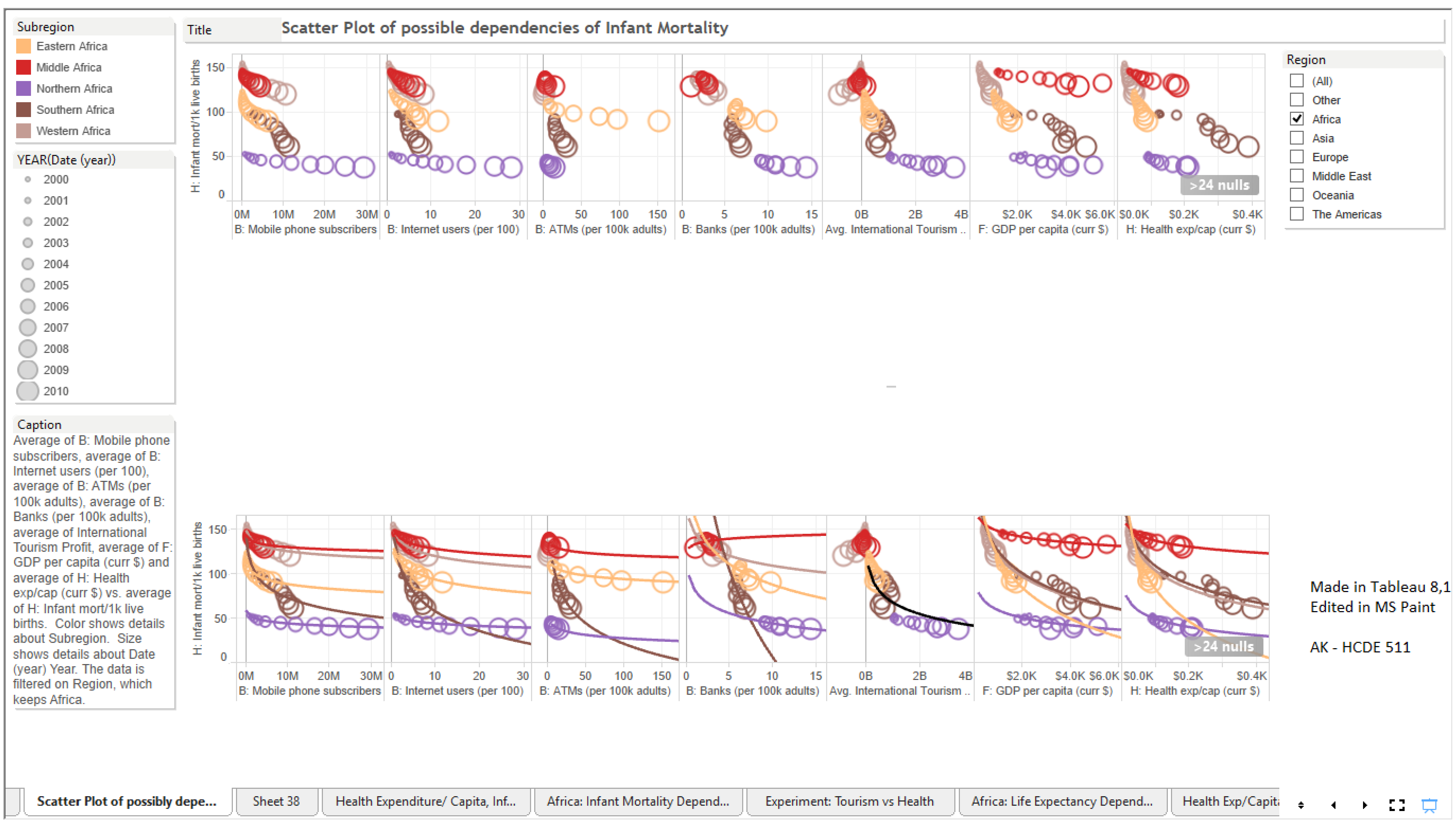

Fig. 17

It would hence be interesting to see in the second iteration if there was any correlation between infant mortality (and possibly life expectancy), and technological & financial parameters, individually.

Iteration 2

In iteration 2, we go forward based on the findings in iteration 1. We have established that we would be focusing on the Health aspect of the African region, and the various measures and dimensions in the Health or other data segments that could affect the quantities under our consideration.

Refined Questions

Is there a correlation between life expectancy and infant mortality in the African Continent?

How do infant mortality and life expectancy correlate with the technology measures in the African continent?

How do infant mortality and life expectancy correlate with the financial measures in the African continent?

How does the life span of males vary against females in the different sub-regions of Africa?

Is there a correlation between life expectancy and infant mortality in the African continent?

[Final Visualization 1 – Fig. 18(a) and Fig. 18(b)]

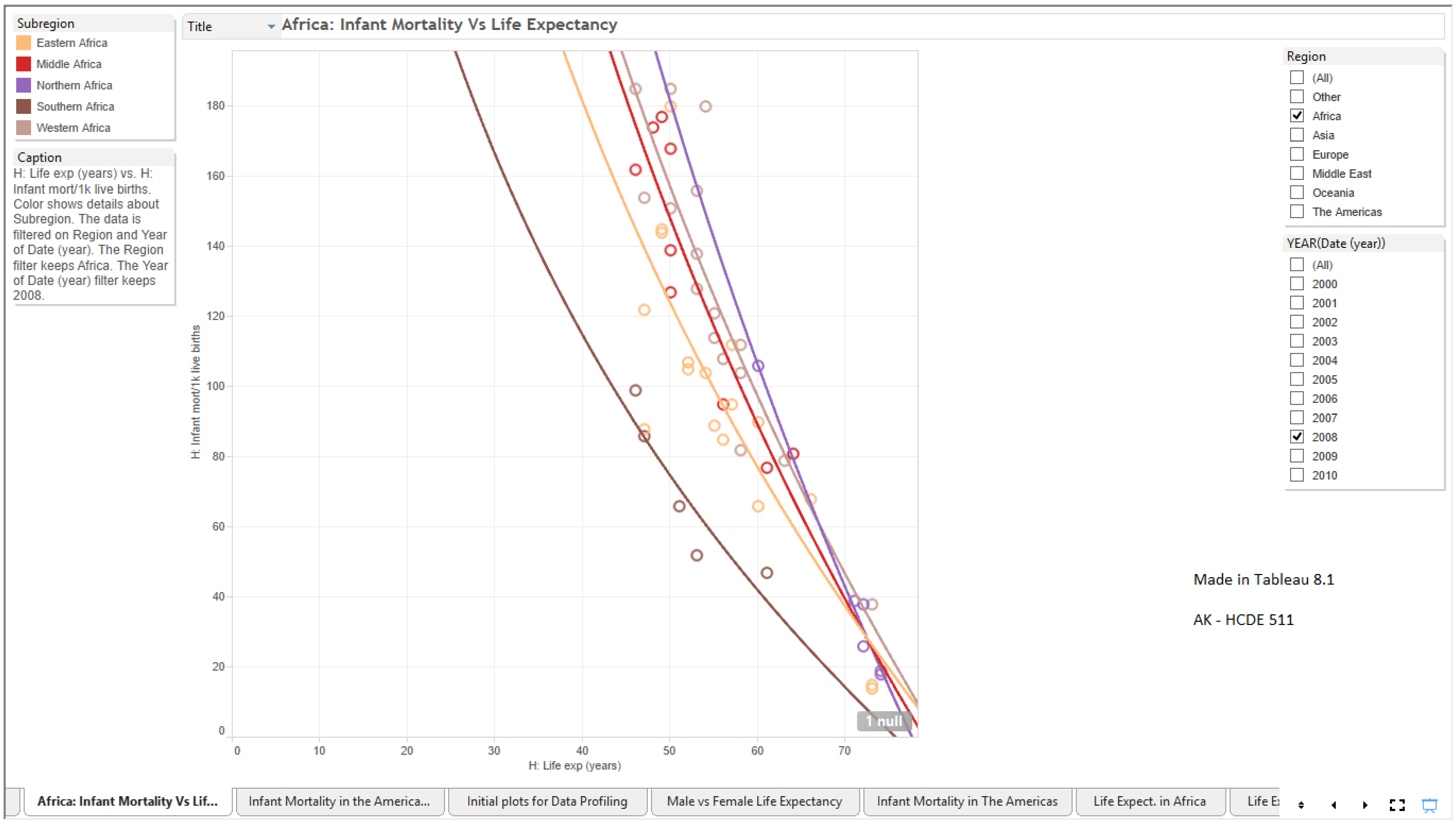

Fig. 18(a)

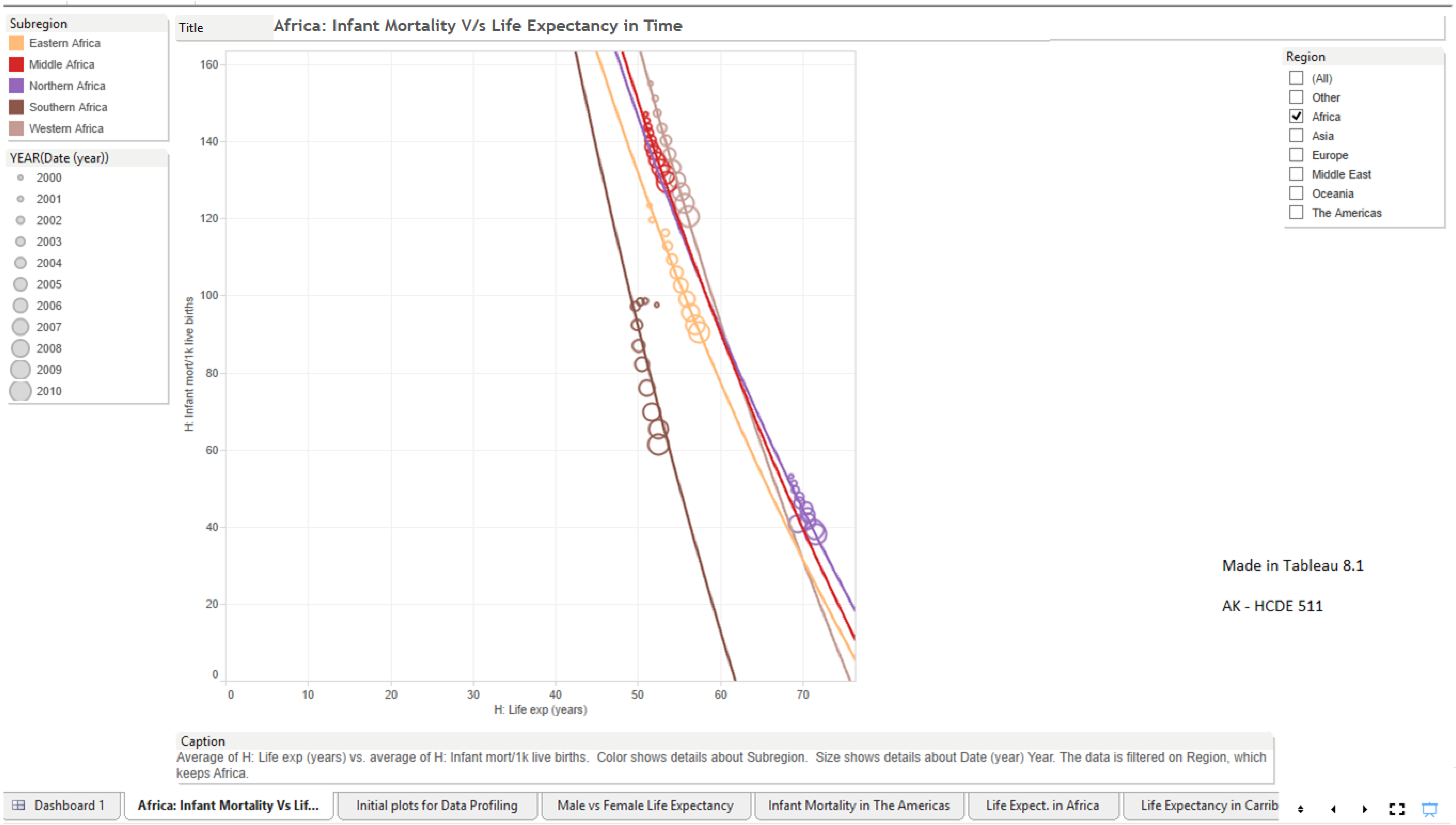

Fig. 18(b)

Description: The above figures (18(a) and 18(b)) are graphs of Infant Mortality versus Life Expectancy for the African continent, which has been color coded as per sub-regions (see legend). The second graph (Fig. 18(b)) also has time encoded in it, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by the sub-region-colored logarithmic lines, pertaining to each sub-region of Africa. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.

As can be seen in Fig. 18(b), there is a very strong inverse correlation between Infant Mortality and Life Expectancy in all the regions of Africa. Not only are these two correlated across the decade, they are also strongly inversely correlated during each year when they are plotted in a non- aggregated manner for any particular year. Fig 18(a) is one such instance of this.

For detailed calculations concerning trend lines model, statistical significance, and correlation (pertaining to Fig. 14(b)), please see **Appendix 1**.

How do infant mortality and life expectancy correlate with the technology measures in the African continent?

Infant Mortality Vs Technology Measures

To start with, here's the global view:

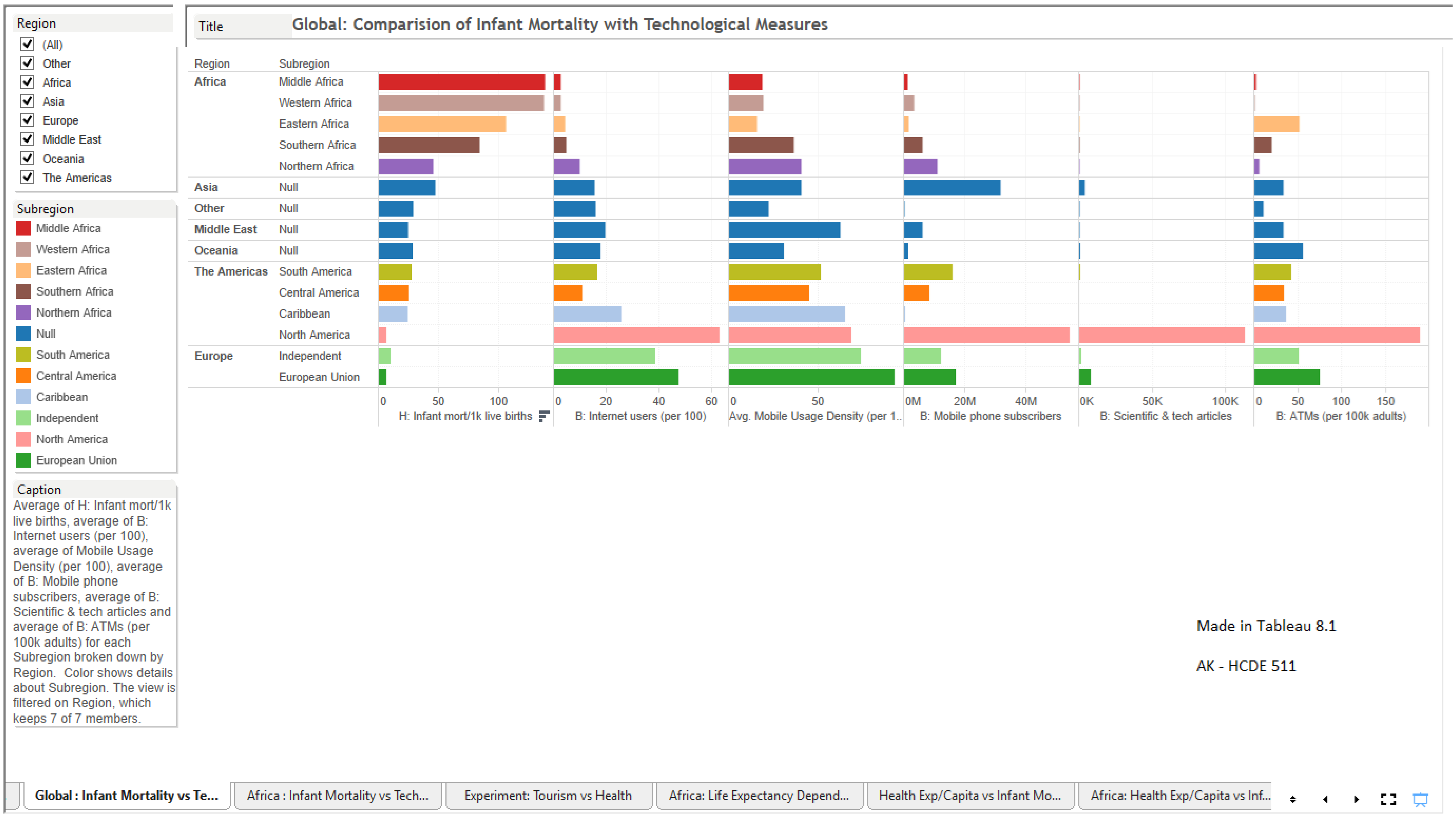

Fig. 19

As can be seen above, there appears to be a strong inverse correlation, particularly between Infant mortality and the number of Internet users and the average mobile usage density.

Drilling down in the direction of our question, we get the following:

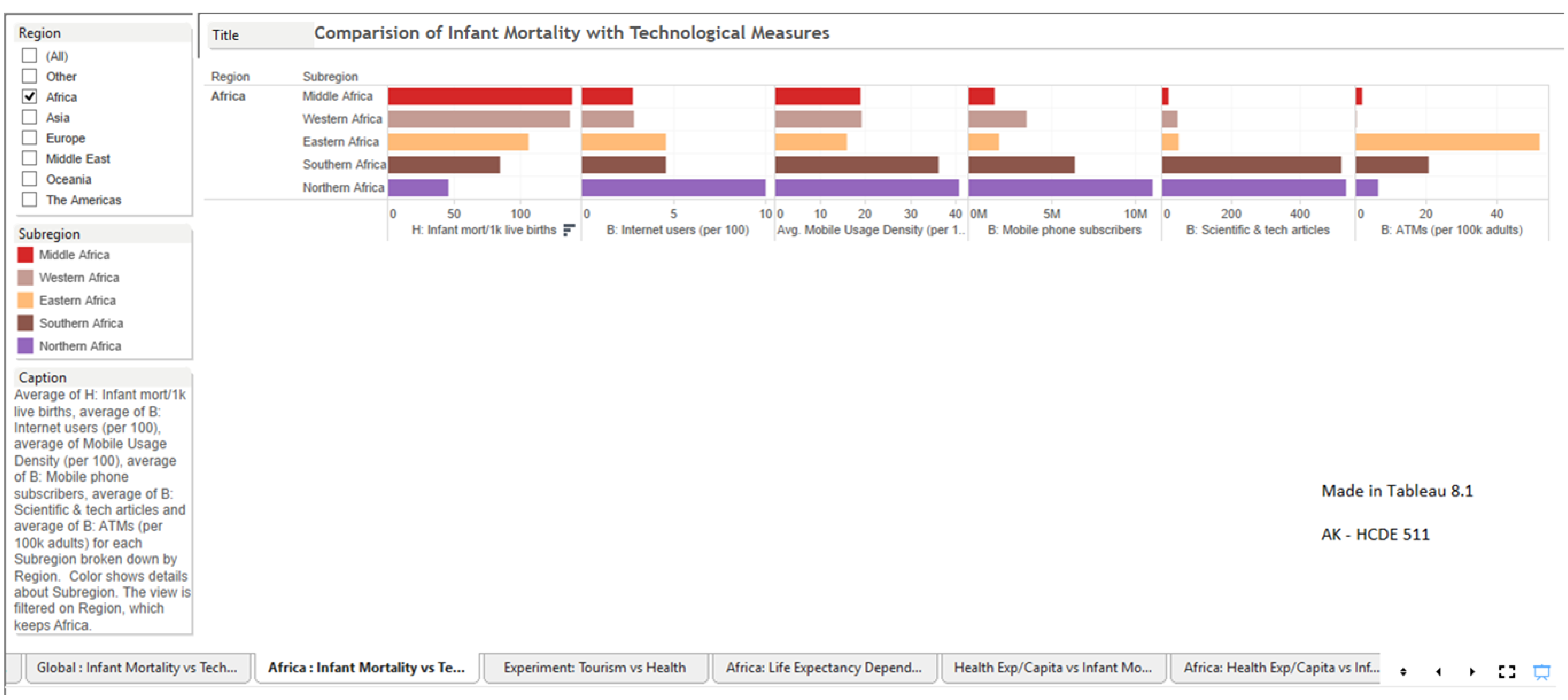

Fig. 20

This is similar to the results generated when analyzing the global view above, except for the number of ATMs (per 100 adults), which does not seems to be correlated (at least visually). Also, since the above graphics do not involve the time scale, the results may need to be looked further into through the time span.

The following scatter plots throws more light on this, and gives stronger conclusions:

[Final Visualization 2 – Fig. 21]

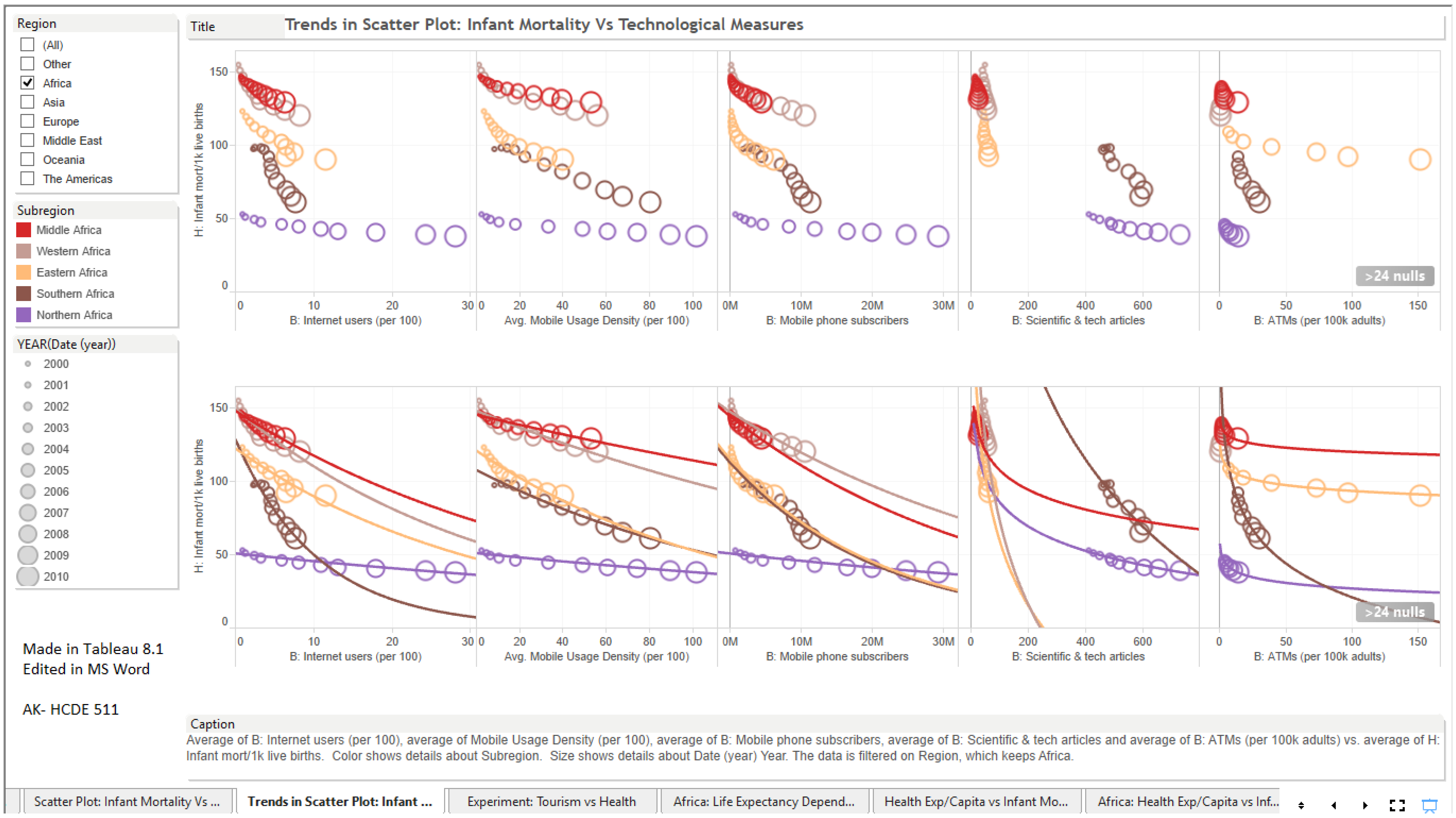

Fig. 21

Description: *The above figure 21 contains graphs concerning Infant Mortality versus the various Technological measures available in the data set for the African continent, which have been color coded as per sub-regions (see legend). The graphs also have time encoded in it, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by the sub-region-colored logarithmic lines, pertaining to each sub-region of Africa. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

As we can see here, this scatter plot, which also accounts for the time scale, clarifies greatly and shows that ALL of these technological measures (including the number of ATMs per 100 adults as well) are strongly correlated (inversely) with infant mortality, thus dispelling our earlier results concerning the correlation between Infant mortality and ATMs.

For detailed calculations concerning trend lines model, statistical significance, and correlation for Fig. 21, please see **Appendix 2(a)**.

Note: One possible reason for the strong inverse correlation between mobile phone usage (and internet usage) and infant mortality is because people in Africa primarily communicate electronically by mobile phones, and access and share information through them. They have a strong network, which has been leveraged by various institutions like University of Washington, Seattle (CSE) (http://opendatakit.org/about/press/li) to improve maternal and neo-natal care in that area. Also, access to these along with other technological parameters mentioned in the graph enables the people to share information and procure resources conveniently (possibly through ATMs), which considerably helps in reducing infant mortality as inadequate access to information (and resources) is one of the main reasons for its existence and rise.

As for Scientific and Tech articles, progress in them is one of the hallmarks of intellectual development of a society, which encourages higher learning and higher access and assimilation of information. As this helps the society to advance intellectually, it would also enable it to sustain itself against infant mortality, which is a direct threat to the survival of a society.

Life Expectancy Vs Technology Measures

Starting on a broader scale:

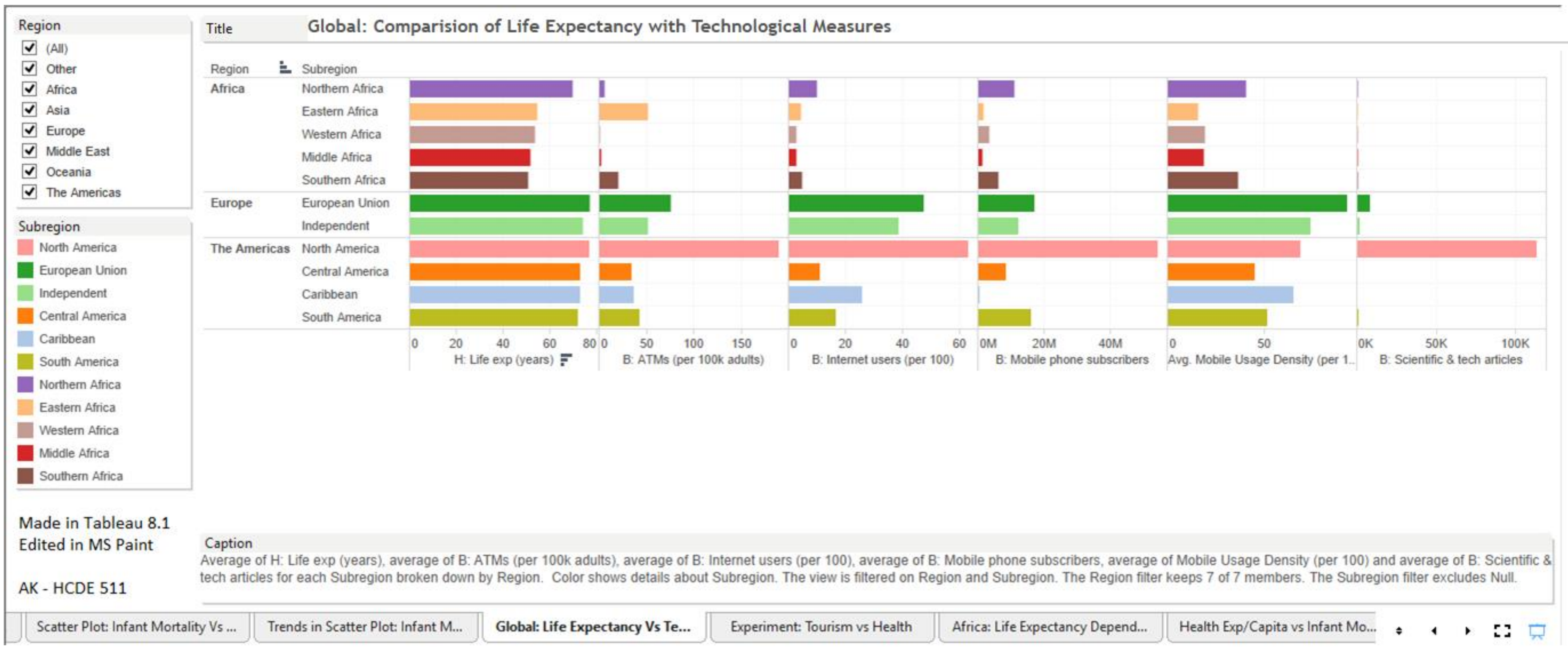

Fig. 22

In this case, there doesn't appear to be a correlation as strong as in the previous case, even though life expectancy and Infant Mortality are inversely correlated.

Drilling down in the direction of our question, we get the following:

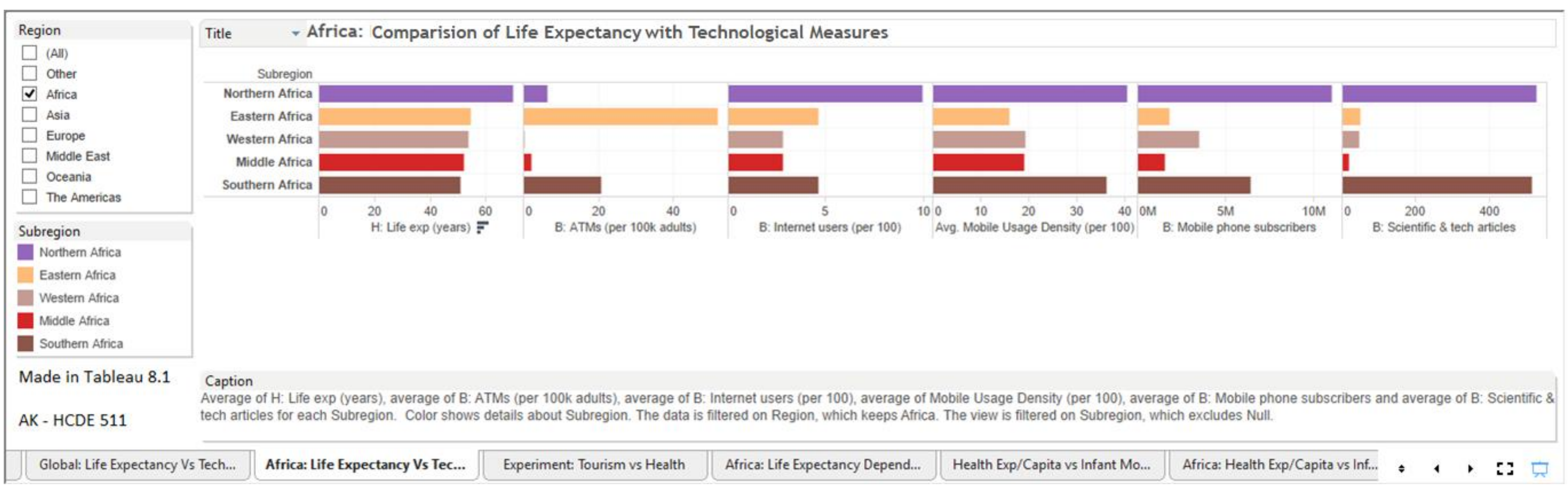

Fig. 23

Even here, it's difficult to decipher a result or a pattern. We see that changing this visualization to stacked bars and adding the time dimension helps considerably in identifying patterns (of related increments) amongst different measures.

[Final Visualization 3 – Fig. 24 and Fig. 25]

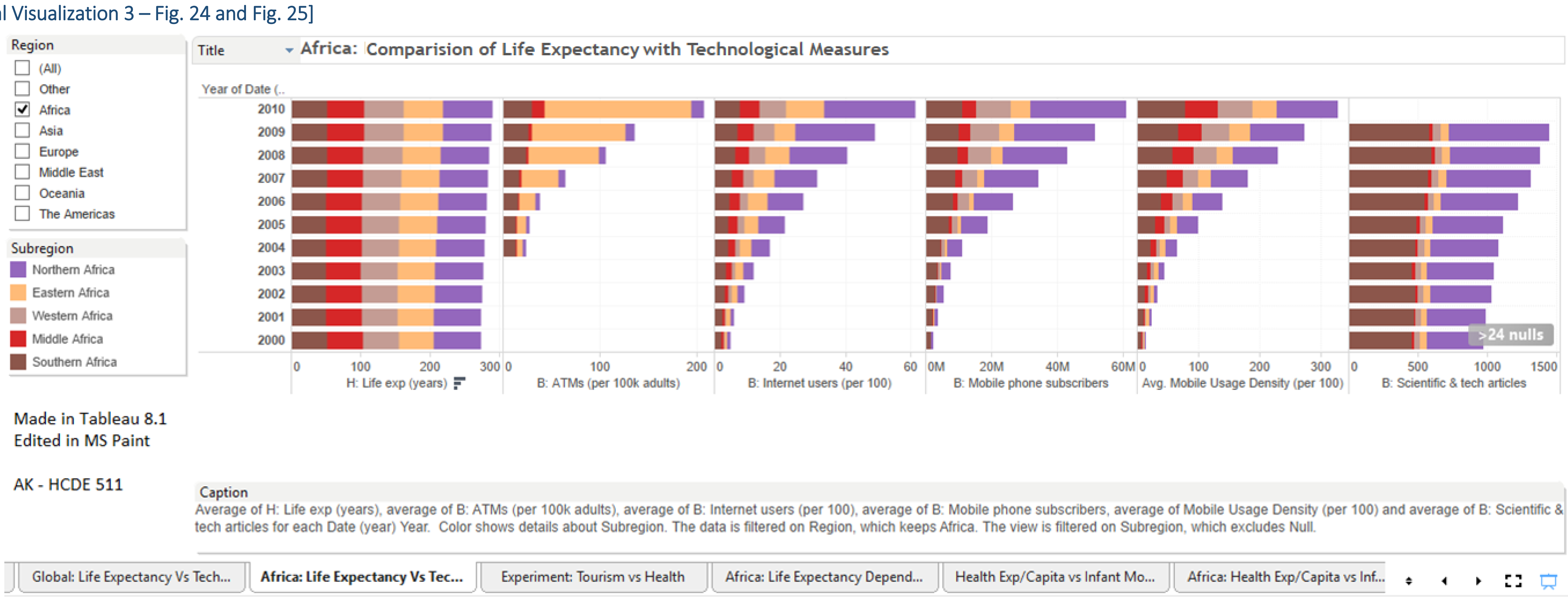

Fig. 24

Description: The above figure 21 contains bar-graphs concerning Time against Life Expectancy and the various Technological measures available in the data set for the African continent, which have been color coded as per sub-regions (see legend). Also, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views that are available.

Now, changing the visualization to a scatterplot with trend lines (to get the hard facts), we have the following (main solution to the question):

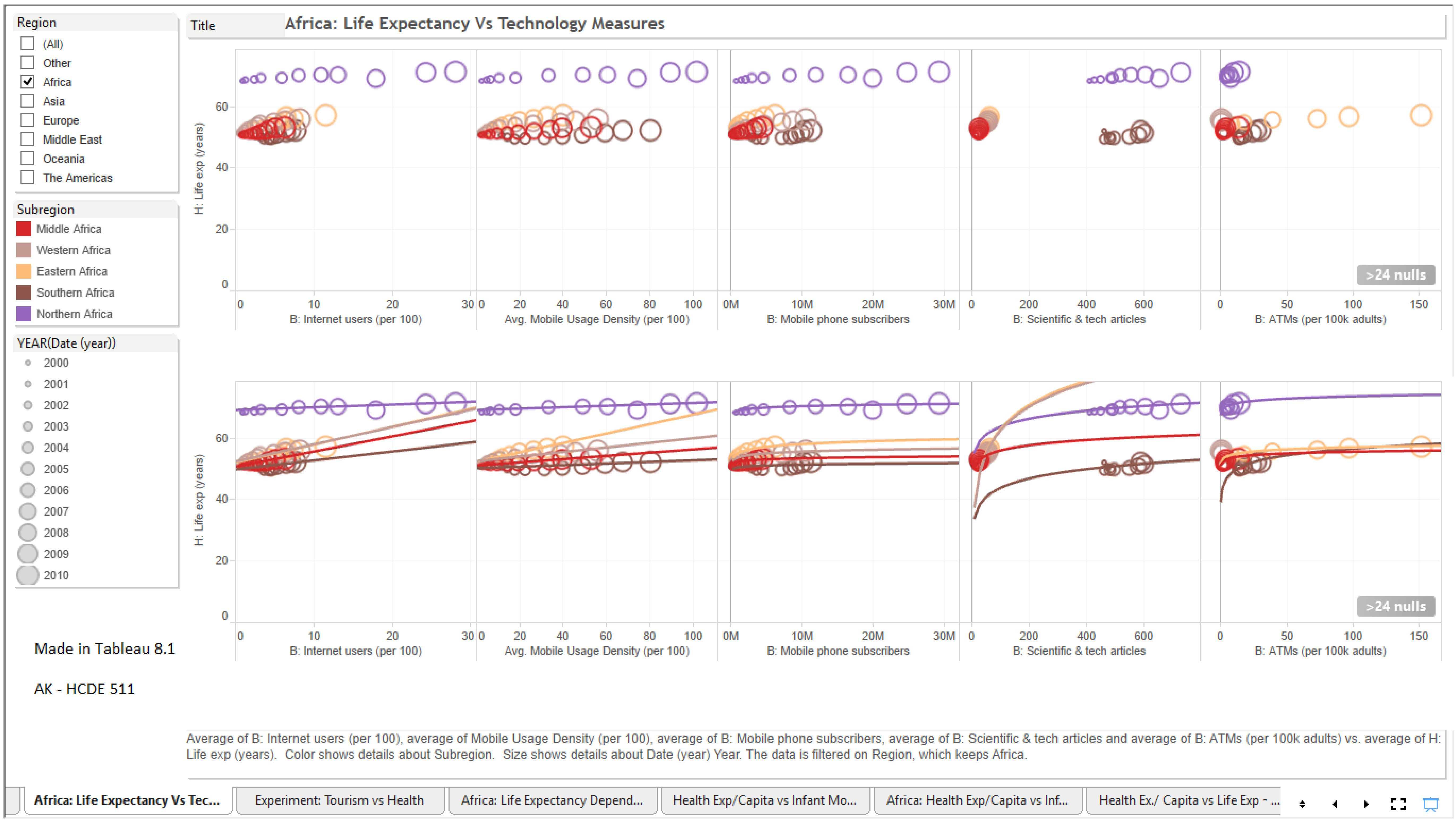

Fig. 25

Description: *The above figure 25 contains graphs concerning Life Expectancy versus the various Technological measures available in the data set for the African continent, which have been color coded as per sub-regions (see legend). The graphs also have time encoded in it, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) are apt. As for trends, these have been depicted by the sub-region-colored logarithmic lines, pertaining to each sub-region of Africa. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

As can be seen here, this greatly enhances comprehension. There is a correlation between Life Expectancy and Technological Measures, though not as strong as in the case of Infant Mortality Vs Technological Measures. (Also, since we have already shown that the infant mortality and life expectancy are strongly correlated with each other, hence in this case life expectancy had to be correlated with the technology measures, as we just showed above that infant mortality was correlated to technological measures as well.)

The main value of this graph is that it depicts that after a certain point, the technological measures (under our consideration) do not effect life expectancy much (or as strongly as they effect infant mortality).

For in-depth documentation of calculations, please refer the **[Appendix 2(b)]** of the document.

How do infant mortality and life expectancy correlate with the financial measures in the African continent?

Infant Mortality Vs Financial Measures

[Final Visualization 4 – Fig. 26 and Fig. 27]

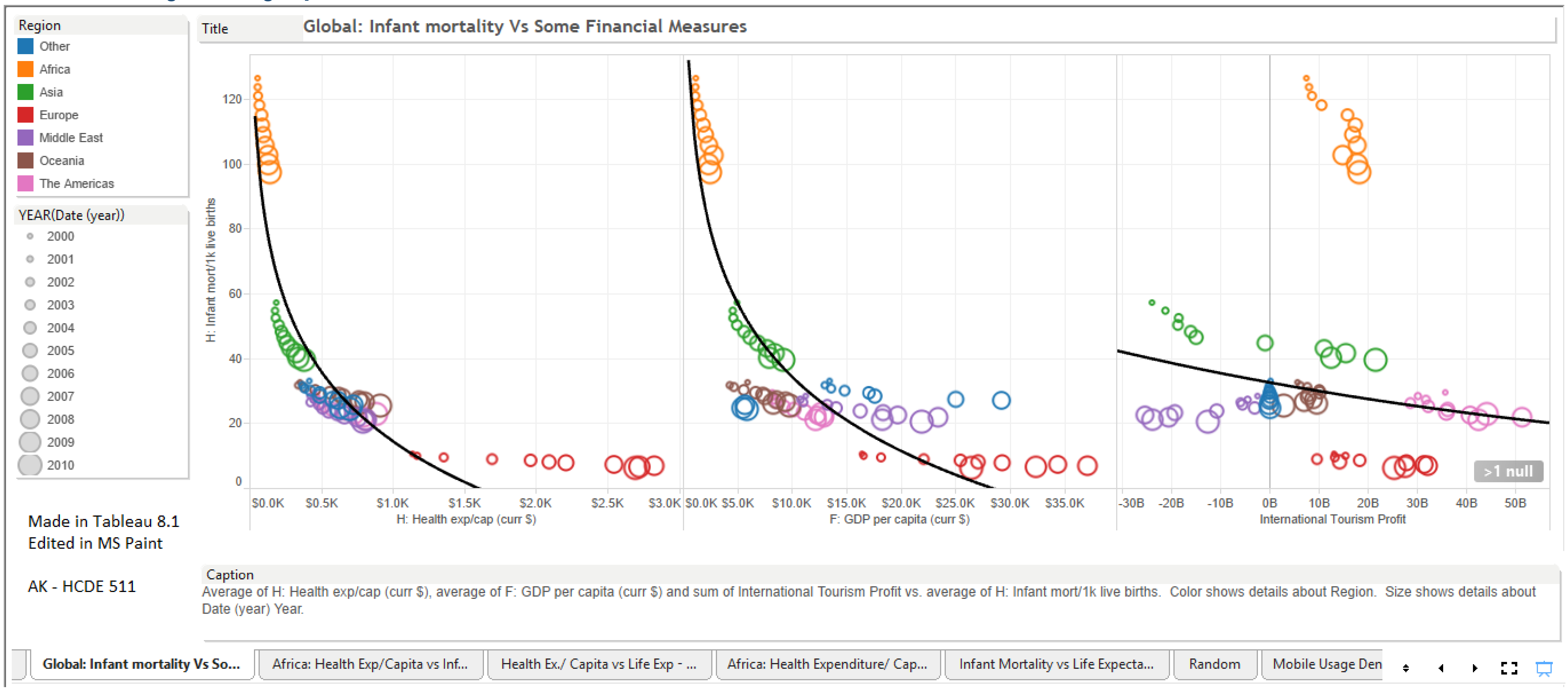

Fig. 26

Description: *The above figure 26 contains graphs concerning Infant Mortality versus the various financial measures available in the data set, which have been color coded as per Regions (see legend). The graphs also have time encoded in it, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by logarithmic and exponential (Infant Mortality V/s International Tourism Profit) lines. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

In the above graph, it is clear that the Health Expenditure per Capita and the GDP per capita are inversely correlated with Infant Mortality. The progression of time encoded by the size of the bubbles also enhances the visual and strengthens this conclusion by the smooth curvature they describe, even on a region specific level. On the other hand, though the graph pertaining to international tourism profit seems to generalize that increase in these profits is inversely correlated with infant mortality, closer inspection shows that almost all of the regions have shown a decline in infant mortality with the progression of time, irrespective of profit or loss.

To gain more clarity pertaining to the African region's context, we drill down to have:

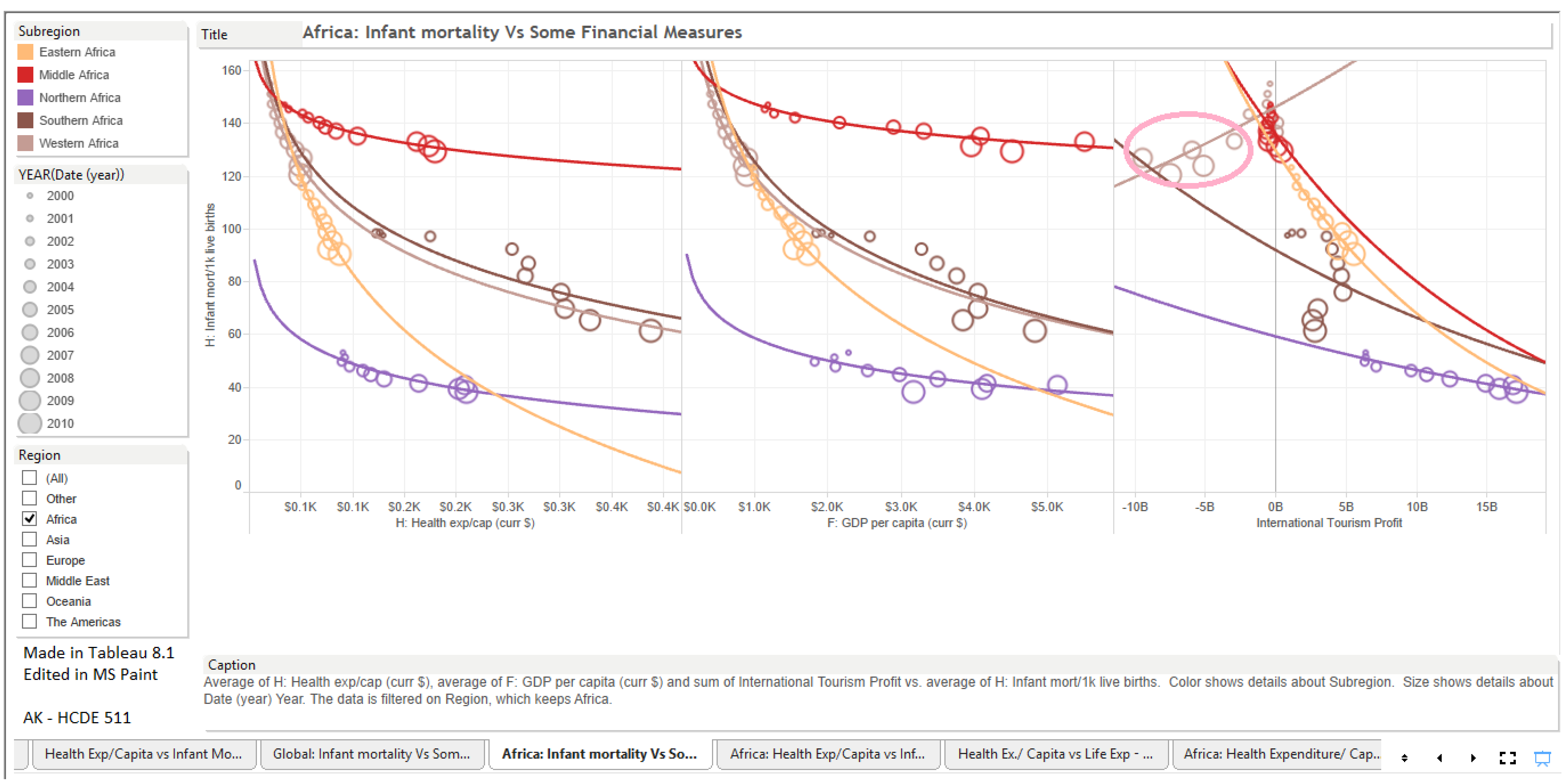

Fig. 27

Description: *The above figure 27 contains graphs concerning Infant Mortality versus the various financial measures available in the data set for the African continent, which have been color coded as per Sub-Regions (see legend). The graphs also have time encoded in them, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by logarithmic and exponential (Infant Mortality V/s International Tourism Profit) lines, which were chosen on the basis of how well they fit the plots and what their scope was. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

As was the case in the global context, the Health Expenditure/Capita and the GDP/Capita show a strong inverse correlation with infant mortality here as well.

However, the third graph pertaining to International tourism profit is interesting. Based on its plot and the trend lines, one could conclude that Tourism, whether in profit or loss (**the Pink oval**), is helpful in lowering Infant Mortality in all the regions of Africa (just as we had felt in the previous graph, which was depicting this at a global level). Perhaps the reason for this could be that tourism improves communication to the outside world and brings in new knowledge and technology with it, which (as shown earlier) have strong inverse correlation with Infant Mortality.

For detailed calculations concerning trend lines model, statistical significance, and correlation, please see Appendix 3(a).

Life Expectancy Vs Some Financial Measures

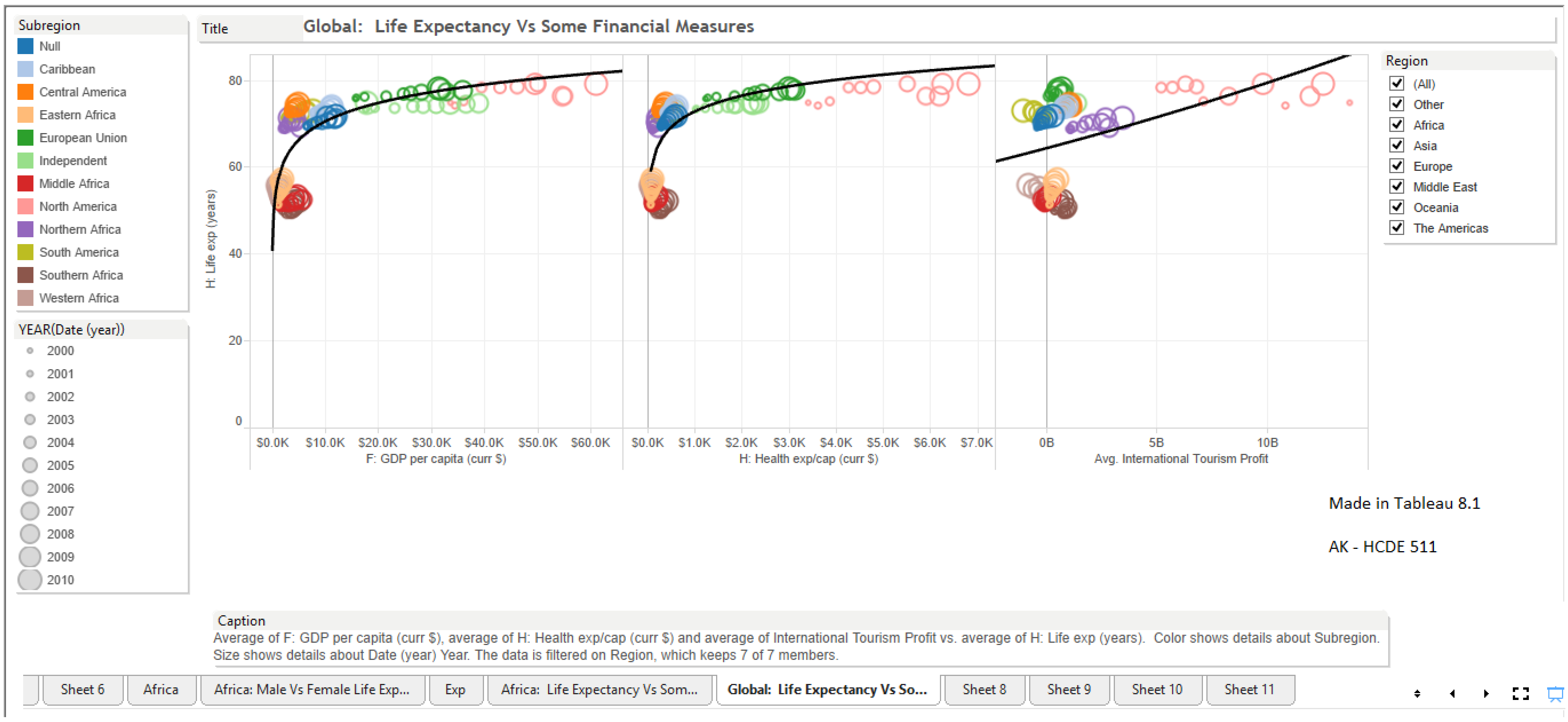

Fig. 28

Description: *The above figure 28 contains graphs concerning Life Expectancy versus the various financial measures available in the data set, which have been color coded as per Regions (see legend). The graphs also have time encoded in it, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by logarithmic and exponential (Life Expectancy V/s International Tourism Profit) lines. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

In the above graph, it is clear that the Health Expenditure per Capita and the GDP per capita are correlated with Life Expectancy. On the other hand, though the graph pertaining to international tourism profit seems to generalize that increase in these profits is correlated with life expectancy, while closer inspection shows that almost all of the regions have shown an increase in life expectancy with the progression of time, irrespective of profit or loss.

To gain more clarity pertaining to the African region's context, we drill down to have:

[Final Visualization 5 – Fig. 29 and Fig. 30]

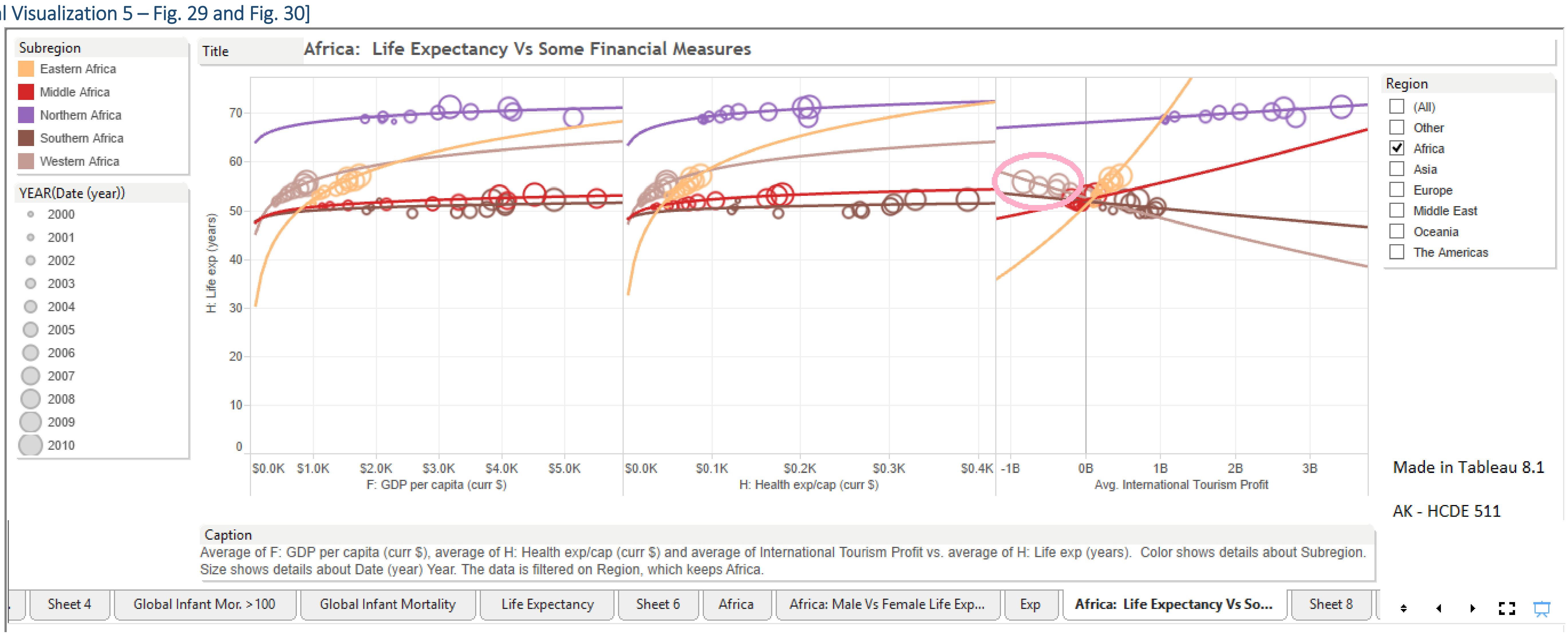

Fig. 29

Description: *The above figure 29 contains graphs concerning Life Expectancy versus the various financial measures available in the data set for the African Region, which have been color coded as per Sub-Regions (see legend). The graphs also have time encoded in it, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by logarithmic and exponential (Life Expectancy V/s International Tourism Profit) lines. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

As was true in the global context, the Health Expenditure/Capita and the GDP/Capita show some correlation with life expectancy.

On the other hand, the third graph pertaining to International tourism profit is interesting. Based on its plot and the trend lines, one could conclude that Tourism, is correlated with an in increase in Life Expectancy in only 3 regions of Africa. This reveals an irregularity as Life Expectancy and Infant Mortality are strongly correlated (inversely), and yet, there are two regions here which do not seem to be adhering to the intuitive conclusion.

The graph below helps in throwing some light on this:

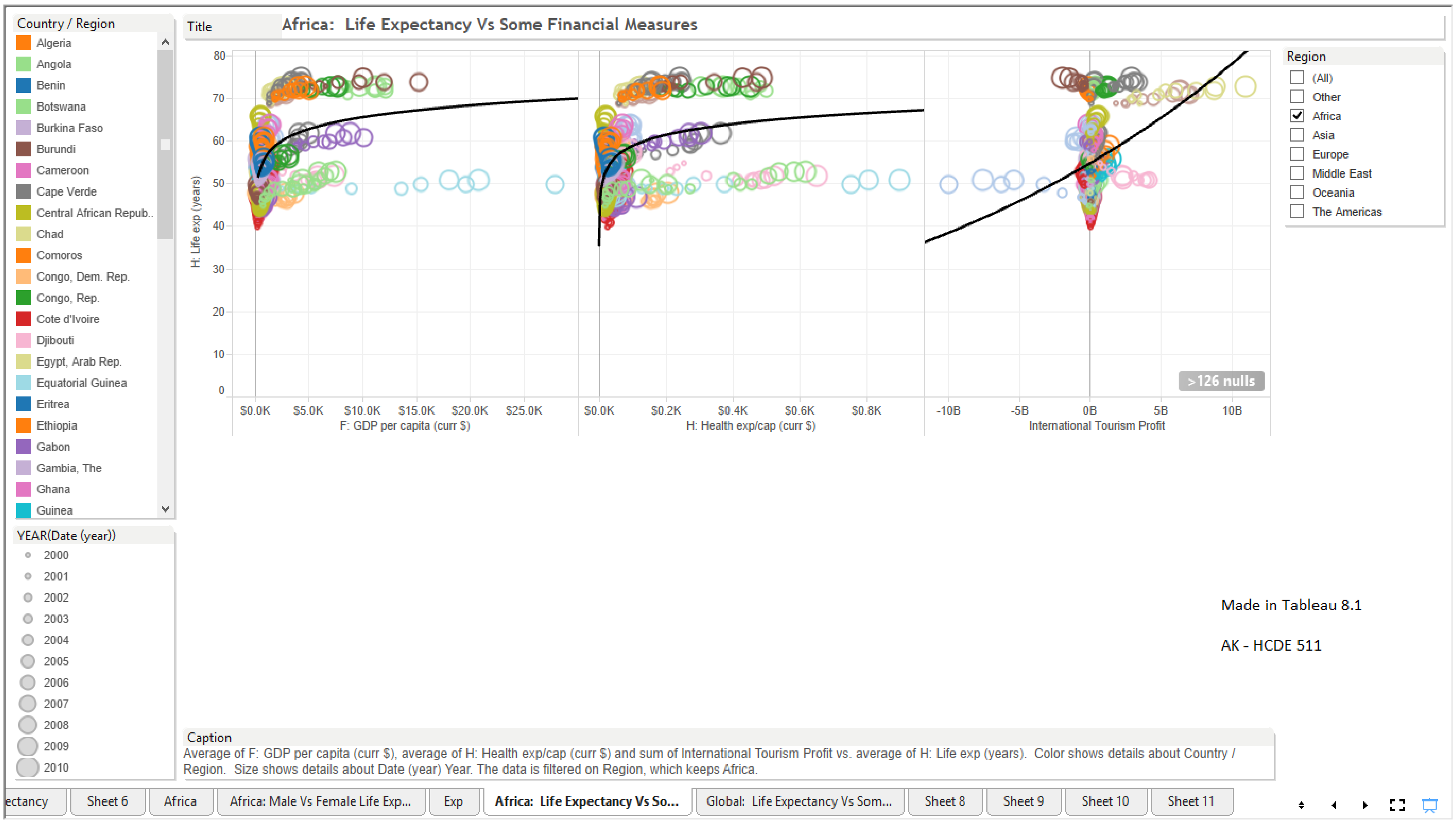

Fig. 30

Description: *The above figure 30 contains graphs concerning Life Expectancy versus the various financial measures available in the data set for the African Region, which have been color coded as per countries (see legend). The graphs also have time encoded in it, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by logarithmic and exponential (Life Expectancy V/s International Tourism Profit) lines. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

From this visualization (Fig. 30), we observe that International Tourism Profit seems to be directly correlated with Life Expectancy. Hence, unlike infant mortality which appeared to decrease irrespective of international tourism being profitable or not, life expectancy (though seemingly strongly correlated (inversely) with infant mortality) seems to be directly correlated with the international tourism's profits.

This could be because while tourism (whether profitable or not) increases access to the outside world in terms of technology and information, if it does not provide adequate income to the masses for their **healthy** sustenance, their life span might be reduced. Also, the income earned is spread across the entire human life span (and/or its different stages) of the individual, while in the case of babies (infants), it would be counted as required only till the time they are in their infancy, which is a much shorter time frame when compared to the entire life span.

(Further, infant Mortality is calculated on the basis of the number of infant deaths per 1000 births. If the infant has grown into a child (no longer infant), and if then something happens to him/her, then this statistic would be counted under life expectancy only, and not under infant mortality.)

For detailed calculations concerning trend lines model, statistical significance, and correlation (concerning Fig.29), please see Appendix 3(b).

How does the life span of males vary against females in the different sub-regions of Africa?

The following graphs address this:

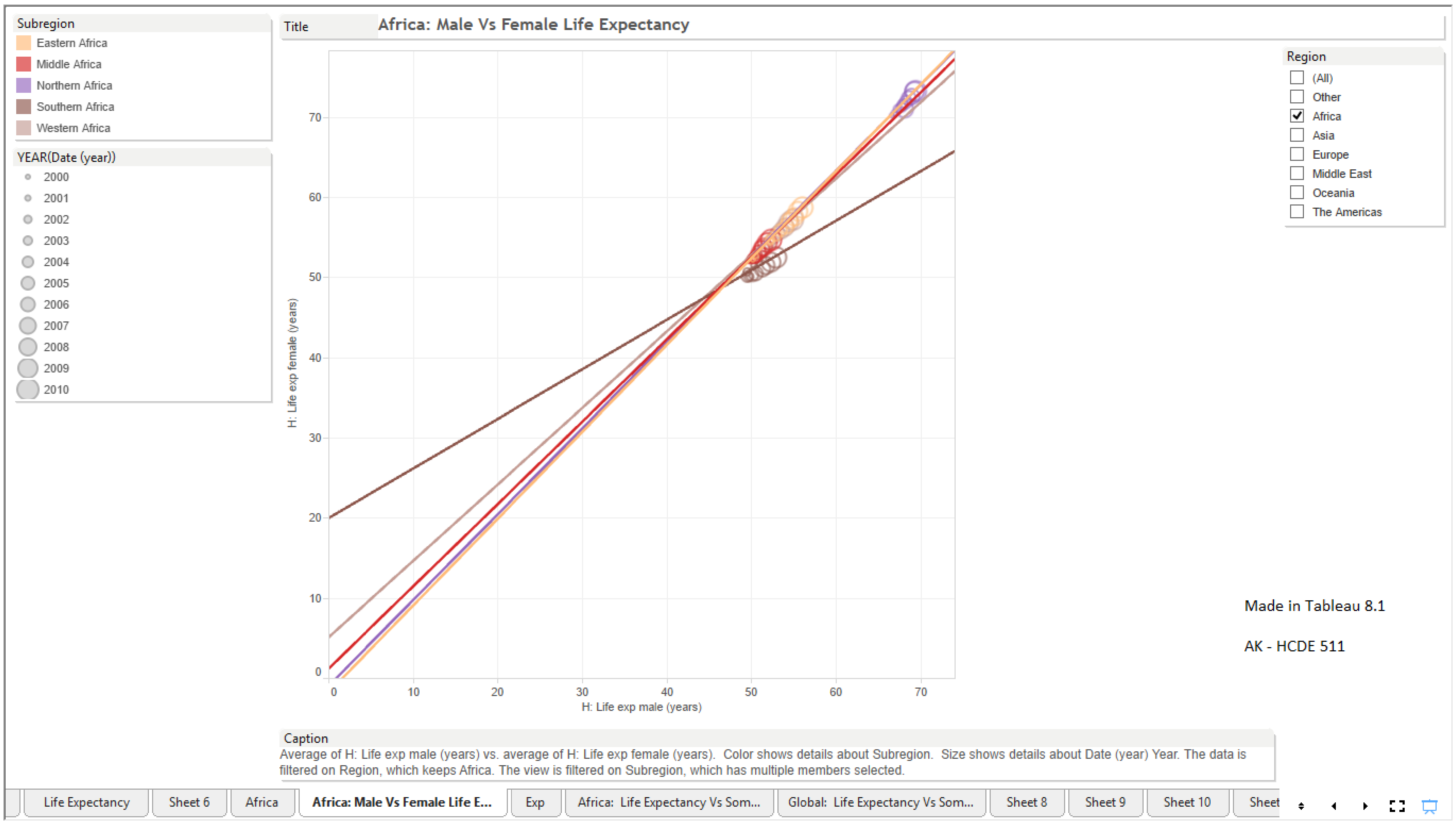

Fig. 31

Description: This is a graph of female versus male life expectancy in the African continent, which has been color coded as per sub-regions. The trend lines are linear in nature and are also color-coded pertaining to each sub-region.

As can be seen in the graph, there is a strong direct correlation between the male and the female life expectancy in the African continent across all the sub-regions.

To gain more insight, we drill further to find:

[Final Visualization 6 – Fig. 32]

Made in Tabeau 8.1 and Edited in MS Paint AK - HCDE 511

Fig. 32

Description: *The above figure 32 contains graphs concerning Male versus Female Life Expectancy available in the data set for the African Region, which have been color coded as per individual countries (see legend) and also have been divided as per sub-regions of Africa. The graphs also have time encoded in it, which is depicted by the size of the circle (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by linear lines. And finally, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

In the above visualization, the five graphs are plotted as per sub-regions, and all of them together have accounted for all the countries in Africa. As can be seen here, all the countries point towards a very strong direct correlation between the male and female life expectancy. However, when we hover on these plots, we find that there is almost always a gap between the female and the male life expectancies, which, with the progression of time, seems to diminish in some places while expand in others. Hypothesizing that this could be pertaining to the standard of living and richness (GDP/capita) of the nation, we employ the brushing and linking approach in the hopes of getting a clearer picture, which is depicted as follows:

[Final Visualization 7 – Fig. 33 and Fig. 34]

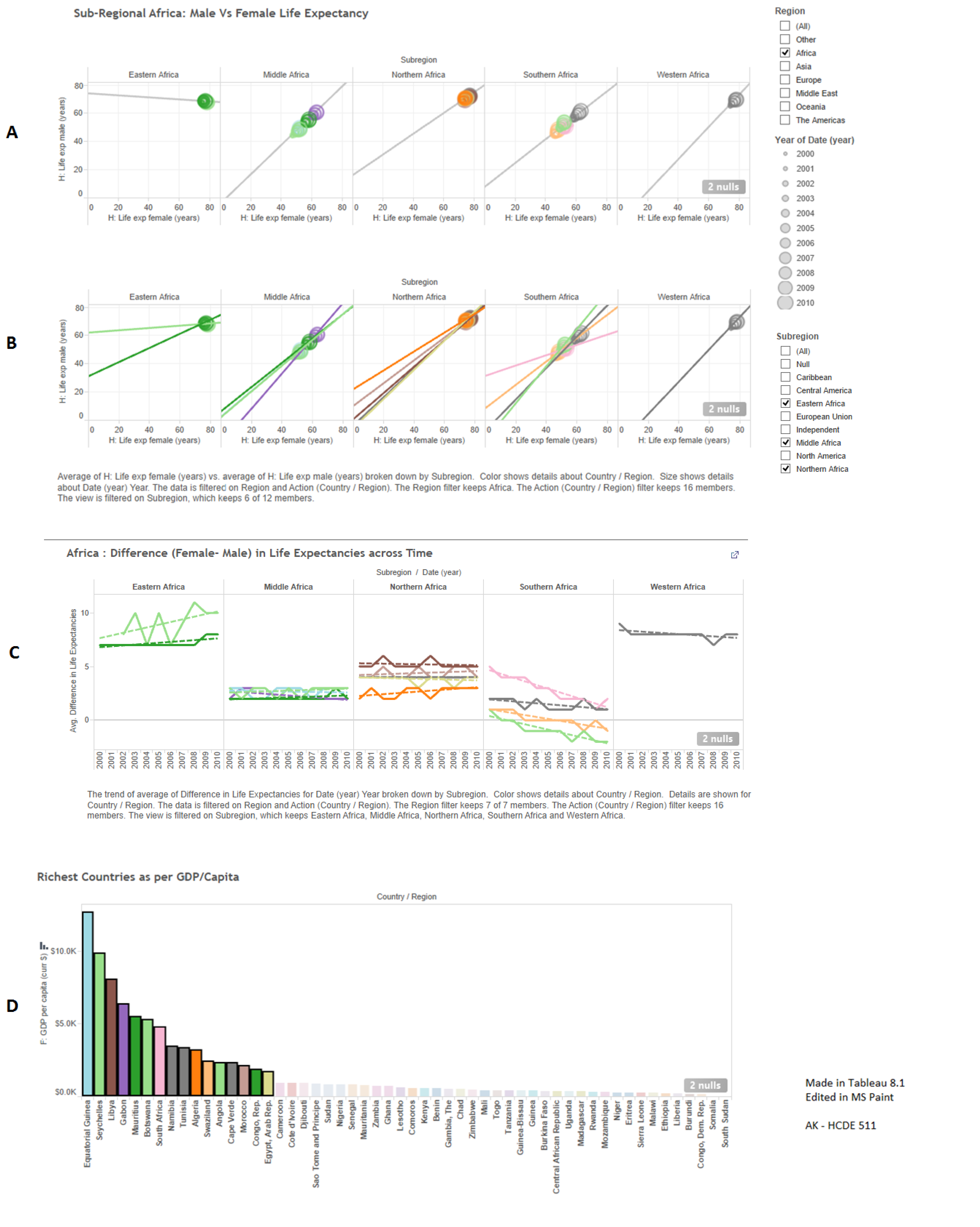

Fig. 33

Description: *The above figure 33 contains graphs concerning Male versus Female Life Expectancy available in the data set for the African Region, which have been color coded as per individual countries (see Richest Countries as per GDP per capita graph) and also have been divided as per sub-regions of Africa. The graphs also have time encoded in it (**A, B, C**), which is depicted by the size of the circle (**in A and B**) (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) is apt. As for trends, these have been depicted by linear continuous lines (**A, B**), and dotted lines (**C**), which have also been colored as per the country they represent. Also, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.*

*And finally, Brushing and Linking has been used here. The countries selected in the last graph (**D**) (Title: Richest Countries as per GDP/Capita) are the ones that are depicted in the first two sets of graphs (Title: Sub-Regional Africa: Male VS Female Life Expectancy). These countries are arranged in descending order of their richness (GDP/capita), and the top rich ones (selected non-empirically on the basis of "keep-going-unless-there's-a-significant-jump-followed-by-plateau-or-further-decrease-in-the-value") have been brought under consideration.*

As can be seen in the above visualization (particularly from graphs **B** and **C**), the female life expectancy is almost always higher than the male life expectancy, except in South Africa, where all the comparatively richer countries seem to moving towards reducing the gap between the two expectancies, while still maintaining correlation.

The main point that can be made here is that the countries which have lower life expectancies in general tend to be moving towards faster increment in male life expectancies in South Africa (which is shown by the difference going towards the negative in graph **C**), which is at odds with what we found in the global scenario.

As for the other sub- regions, the difference in the life expectancies between females and males has either been increasing, or has remained almost the same, despite higher per capita income. This, too, is at odds with the global scenario, wherein we saw that in the beginning, females tend to have higher life expectancies than males. But as the nations become progressively wealthy, these differences start to diminish.

Now, to gauge the scenario in the rest of the countries (the comparatively poorer ones) in the African continent, we have the following:

Fig. 34

Description: The above figure 34 contains graphs concerning Male versus Female Life Expectancy available in the data set for the African Region, which have been color coded as per individual countries (see Richest Countries as per GDP per capita graph) and also have been divided as per sub-regions of Africa. The graphs also have time encoded in it (A, B, C), which is depicted by the size of the circle (in A and B) (greater the size, more recent is the data point). Such an encoding does the job of depicting the progression of years remarkably well, and since the main purpose of the graph is to find correlation and trends, this type of encoding (along with others in this graph) are apt. As for trends, these have been depicted by linear continuous lines (A, B), and dotted lines (C), which have also been colored as per the country they represent. Also, the various cards (boxes) surrounding the graph provide metadata about the graph, and when dynamic, some of them serve as filters (e.g. Region) and enable us to see different views as available.

*And finally, Brushing and Linking has been used here. The countries selected in the last graph (**D**) (Title: Richest Countries as per GDP/Capita) are the ones that are depicted in the first two sets of graphs (Title: Sub-Regional Africa: Male VS Female Life Expectancy). These countries are arranged in descending order of their richness (GDP/capita), and the remaining comparatively poorer ones are brought into consideration here.*

In this case, females tend to live longer than men, and the gap between their life expectancies seems to change depending upon their actual life expectancies (the gaps seem to be decreasing between the ones with comparatively higher life expectancies, and vice versa). This confirms to the global scenarios that we discussed earlier.

Also, **Lesotho (Southern Africa)** and Zimbabwe (Eastern Africa) are outliers here as even though both of them have comparatively lower life expectancies, the male life expectancy seems to be outgrowing the female life expectancy, and the gap between the two has been increasing progressively with the advent of time. This is in contrast with the discussion we had on the global scale, according to which the female life expectancy should have been greater than the male life expectancy.

To conclude from Fig. 33 and Fig. 34, life expectancies in Africa exhibit different traits when compared to a global scale of things. In most general terms, the female population tends to live longer than the male population, irrespective of the prosperity of the nation they belong to. One good reason behind this could be that the life expectancy itself in the African continent is less when compared to the global scale. Hence, it will take a while in Africa before the female life expectancy reaches its optimum maturity and the male life expectancy starts catching up and causes a reduction in the gaps between them. (Please refer to the discussion on the relationship between male and female life expectancy in Iteration 1).

Features of Tableau

Positives

1. It's really easy to form new calculated variables. One can do that by the sole use of a mouse, which would make this as an excellent tool for those who do not wish to type down the commands.
2. Its feature of disabling the graphs that wouldn't be appropriate at all is very good in determining and implementing this.
3. The ease with which the rows and columns can be changed through just drag and drop is really user friendly. Also, the reaction time, even when dealing with complex calculations is very fast.
4. It affords the concept of minimalistic design to the user – the things that are not required can be removed very easily from the screen. Also, if they are required back, it can be done easily quite well as the use case for these is very intuitive.
5. Supports high modularity. Users can remove (and add) very specific types of things from the display without effecting other things.
6. The dashboard feature is excellent to gain a holistic view of different graphs and perform brushing and linking operations.
7. Extremely well structured with well thought out restrictions and abilities – excellent software for both beginners and advanced users.

Possible Improvements

1. Allowing the user a higher degree of control in manipulating the visualization in terms of color, size, etc. would be useful. For example, in a graph, or any visualization, generally only 1 type of variable (dimension, measure) can be color coded.
2. Tableau operates on the best visualization principles. While this is a good thing, these principles are generalized concepts that work most of the time, but in some instances (as aforementioned), having some flexibility with these rules would help in exploring new ways and aid creativity.
3. It would be useful if Tableau allows us to import our own set of Marks.
4. Regarding trend lines: It would be great if we can edit trend lines on a much lower level – as in creating functions for fitting the distribution, or having the option of tweaking the logarithmic and exponential functions to achieve the same.
5. Further, given a distribution, the software should have the ability to identify the trend line function type (log, exp, poly, etc.) that will **appropriately** fit it.
6. Having the ability to move the graphs around by simple hold and drag would be awesome for creating customized presentations of multiple graphs. For example, graphs are arranged in line or one below the other depending on the number of variables added to the columns and rows. Though this is useful most of the time, sometimes it would be good to have the ability to change the positioning of these graphs on the visualization pane as this could afford easier comparisons.
7. Options to evaluate a graphic on a few usability parameters. For example, it would be great if the software can tell if a graph's color coding would be useless or change meaning for somebody who has color blindness. (As Few mentions in his book "*Now You See It*", color blind people usually have difficulty in differentiating red and green from each other.)
8. More targeted/ specific reports on errors, incompleteness, null values, in the data.
9. A timeline feature to track analysis/ progress (something like version control) would be infinitely helpful. Developing this would be beneficial for Tableau as well as the statistic collected on this could help Tableau to refine its UI further.
10. Improve panning and zooming features. Also, would recommend adding the ability to grab and move around to different parts of a zoomed-in graph with mouse/ finger (if touch is enabled).
11. Would recommend incorporating the ability to paint/mark in terms of another. For example, marking the plots of different countries using color, which depends upon the respective country's GDP.
12. The software should use efficient saving algorithm to occupy lesser space in the memory (Maybe use versioning, and save only the changes compared to the base or prior visualization). The computer starts throwing warnings when the consumption reaches around 500 MB in the memory.
13. It would be great if the interface can be made more responsive to the changes in the size of windows and panes.
14. It would be great if the presentation mode would enable manual re-alignment and edits.

References/ Works Cited

1. Few, S. (2009), *Now you see it: Simple visualization techniques for quantitative analysis.* Oakland, Califonrnia: Analytics Press
2. Life Expectancy. (n.d.). Retrieved April 21, 2014 from the Wikipedia: https://en.wikipedia.org/wiki/Life_expectancy

Appendices

Appendix 1 | Back

A logarithmic trend model is computed for average of H: Infant mort/1k live births given average of H: Life exp (years). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(log(H: Life exp (years)) + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45
SSE (sum squared error):	1281.64
MSE (mean squared error):	28.481
R-Squared:	0.982389
Standard error:	5.33676
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	29385.478	3673.18	128.97	< 0.0001

Individual trend lines:

Panes Row	Column	Color Subregion	Line p-value	DF	Coefficients Term	Value	StdErr	t-value	p-value
H: Infant mort/1k live births	H: Life exp (years)	Western Africa	< 0.0001	9	log(H: Life exp (years))	-399.524	5.96078	-67.0254	< 0.0001
					intercept	1729.09	23.7482	72.8095	< 0.0001
H: Infant mort/1k live births	H: Life exp (years)	Southern Africa	0.0350342	9	log(H: Life exp (years))	-434.48	175.246	-2.47926	0.0350342
					intercept	1792.13	688.824	2.60173	0.028657
H: Infant mort/1k live births	H: Life exp (years)	Northern Africa	0.0004758	9	log(H: Life exp (years))	-303.204	56.906	-5.32814	0.0004758
					intercept	1332.92	241.671	5.51545	0.0003727
H: Infant mort/1k live births	H: Life exp (years)	Middle Africa	< 0.0001	9	log(H: Life exp (years))	-330.397	25.4164	-12.9994	< 0.0001
					intercept	1443.42	100.364	14.3818	< 0.0001
H: Infant mort/1k live births	H: Life exp (years)	Eastern Africa	< 0.0001	9	log(H: Life exp (years))	-298.426	13.4665	-22.1606	< 0.0001
					intercept	1299.56	53.844	24.1356	< 0.0001

Appendix 2(a) | Back

Trend Lines Model

A logarithmic trend model is computed for average of H: Infant mort/1k live births given average of B: ATMs (per 100k adults). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(log(B: ATMs (per 100k adults)) + intercept)
Number of modeled observations:	28
Number of filtered observations:	27
Model degrees of freedom:	8
Residual degrees of freedom (DF):	20
SSE (sum squared error):	55.7523
MSE (mean squared error):	2.78761
R-Squared:	0.998332
Standard error:	1.66961
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	6	29960.973	4993.5	1791.32	< 0.0001

An exponential trend model is computed for natural log of average of H: Infant mort/1k live births given average of B: Internet users (per 100). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(B: Internet users (per 100) + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45
SSE (sum squared error):	0.0503087
MSE (mean squared error):	0.001118
R-Squared:	0.99503
Standard error:	0.0334361
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	5.8054864	0.725686	649.109	< 0.0001

An exponential trend model is computed for natural log of average of H: Infant mort/1k live births given average of B: Mobile phone subscribers. The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(B: Mobile phone subscribers + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45

SSE (sum squared error): 0.063559
MSE (mean squared error): 0.0014124
R-Squared: 0.993721
Standard error: 0.0375822
p-value (significance): < 0.0001
Analysis of Variance:

Strong Correlation

Field	DF	SSE	MSE	F	p-value
Subregion	8	5.3494609	0.668683	473.43	< 0.0001

An exponential trend model is computed for natural log of average of H: Infant mort/1k live births given average of Mobile Usage Density (per 100). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula: Subregion*(Avg. Mobile Usage Density (per 100) + intercept)
Number of modeled observations: 55
Number of filtered observations: 0
Model degrees of freedom: 10
Residual degrees of freedom (DF): 45
SSE (sum squared error): 0.0289491
MSE (mean squared error): 0.0006433
R-Squared: 0.99714
Standard error: 0.0253636
p-value (significance): < 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	6.998156	0.87477	1359.79	< 0.0001

A logarithmic trend model is computed for average of H: Infant mort/1k live births given average of B: Scientific & tech articles. The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula: Subregion*(log(B: Scientific & tech articles) + intercept)
Number of modeled observations: 50
Number of filtered observations: 5
Model degrees of freedom: 10
Residual degrees of freedom (DF): 40
SSE (sum squared error): 717.333
MSE (mean squared error): 17.9333
R-Squared: 0.989079
Standard error: 4.23477
p-value (significance): < 0.0001

Strong Correlation

Individual trend lines:

Panes		Color	Line		Coefficients				
Row	**Column**	**Subregion**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
H: Infant mort/1k live births	B: Internet users (per 100)	Northern Africa	< 0.0001	9	B: Internet users (per 100)	-0.0111428	0.0012174	-9.15261	< 0.0001
					intercept	3.92387	0.0167797	233.846	< 0.0001
H: Infant mort/1k live births	B: Internet users (per 100)	Southern Africa	< 0.0001	9	B: Internet users (per 100)	-0.0949549	0.0078755	-12.0571	< 0.0001
					intercept	4.86035	0.0387799	125.331	< 0.0001
H: Infant mort/1k live births	B: Internet users (per 100)	Eastern Africa	< 0.0001	9	B: Internet users (per 100)	-0.0311485	0.0035737	-8.71603	< 0.0001
					intercept	4.80729	0.0199103	241.448	< 0.0001
H: Infant mort/1k live births	B: Internet users (per 100)	Western Africa	< 0.0001	9	B: Internet users (per 100)	-0.0304983	0.0038029	-8.01986	< 0.0001
					intercept	5.00675	0.0142413	351.565	< 0.0001
H: Infant mort/1k live births	B: Internet users (per 100)	Middle Africa	< 0.0001	9	B: Internet users (per 100)	-0.0231728	0.001201	-19.2952	< 0.0001
					intercept	4.99699	0.0039519	1264.44	< 0.0001
H: Infant mort/1k live births	Mobile Usage Density (per 100)	Northern Africa	< 0.0001	9	Avg. Mobile Usage Density (per 100)	-0.0029449	0.00025	-11.7794	< 0.0001
					intercept	3.92697	0.0133467	294.228	< 0.0001
H: Infant mort/1k live births	Mobile Usage Density (per 100)	Southern Africa	< 0.0001	9	Avg. Mobile Usage Density (per 100)	-0.0070788	0.0002342	-30.2222	< 0.0001
					intercept	4.6782	0.0101128	462.603	< 0.0001
H: Infant mort/1k live births	Mobile Usage Density (per 100)	Eastern Africa	< 0.0001	9	Avg. Mobile Usage Density (per 100)	-0.0082798	0.0008551	-9.68234	< 0.0001
					intercept	4.79357	0.0168138	285.098	< 0.0001
H: Infant mort/1k live births	Mobile Usage Density (per 100)	Western Africa	< 0.0001	9	Avg. Mobile Usage Density (per 100)	-0.0040212	0.00048	-8.37833	< 0.0001
					intercept	4.99692	0.0127927	390.607	< 0.0001
H: Infant mort/1k live births	Mobile Usage Density (per 100)	Middle Africa	< 0.0001	9	Avg. Mobile Usage Density (per 100)	-0.002411	0.0001992	-12.1036	< 0.0001
					intercept	4.97781	0.0049783	999.899	< 0.0001
H: Infant mort/1k live births	B: Mobile phone subscribers	Northern Africa	< 0.0001	9	B: Mobile phone subscribers	-1.045e-008	1.035e-009	-10.089	< 0.0001
					intercept	3.92259	0.0151823	258.365	< 0.0001
H: Infant mort/1k live births	B: Mobile phone subscribers	Southern Africa	< 0.0001	9	B: Mobile phone subscribers	-4.77e-008	5.05e-009	-9.44622	< 0.0001
					intercept	4.727	0.0364453	129.701	< 0.0001
H: Infant mort/1k live births	B: Mobile phone subscribers	Eastern Africa	< 0.0001	9	B: Mobile phone subscribers	-4.701e-008	6.648e-009	-7.0723	< 0.0001
					intercept	4.74821	0.0178702	265.705	< 0.0001
H: Infant mort/1k live births	B: Mobile phone subscribers	Western Africa	< 0.0001	9	B: Mobile phone subscribers	-2.1e-008	2.439e-009	-8.60905	< 0.0001
					intercept	4.99319	0.0121601	410.62	< 0.0001
H: Infant mort/1k live births	B: Mobile phone subscribers	Middle Africa	< 0.0001	9	B: Mobile phone subscribers	-2.653e-008	2.343e-009	-11.3239	< 0.0001
					intercept	4.97293	0.0049909	996.395	< 0.0001
H: Infant mort/1k live births	B: Scientific & tech articles	Northern Africa	< 0.0001	8	log(B: Scientific & tech articles)	-24.1809	1.98516	-12.1808	< 0.0001
					intercept	197.271	12.4266	15.8749	< 0.0001
H: Infant mort/1k live births	B: Scientific & tech articles	Southern Africa	< 0.0001	8	log(B: Scientific & tech articles)	-113.509	11.6084	-9.77822	< 0.0001
					intercept	795.134	72.4637	10.9729	< 0.0001
H: Infant mort/1k live births	B: Scientific & tech articles	Eastern Africa	0.0007259	8	log(B: Scientific & tech articles)	-65.1011	12.2766	-5.30286	0.0007259
					intercept	360.363	47.6273	7.56632	< 0.0001
H: Infant mort/1k live births	B: Scientific & tech articles	Western Africa	0.0042559	8	log(B: Scientific & tech articles)	-84.1857	21.3309	-3.94666	0.0042559
					intercept	462.525	81.9841	5.64164	0.0004861
H: Infant mort/1k live births	B: Scientific & tech articles	Middle Africa	0.0003254	8	log(B: Scientific & tech articles)	-19.5562	3.26236	-5.9945	0.0003254
					intercept	197.703	9.71087	20.3589	< 0.0001
H: Infant mort/1k live births	B: ATMs (per 100k adults)	Northern Africa	< 0.0001	5	log(B: ATMs (per 100k adults))	-5.53759	0.442159	-12.524	< 0.0001
					intercept	52.1417	0.822726	63.3768	< 0.0001
H: Infant mort/1k live births	B: ATMs (per 100k adults)	Southern Africa	0.0001996	5	log(B: ATMs (per 100k adults))	-33.4266	3.45251	-9.68181	0.0001996

					intercept	174.65	10.1924	17.1353	< 0.0001
H: Infant mort/1k live births	B: ATMs (per 100k adults)	Eastern Africa	< 0.0001	5	log(B: ATMs (per 100k adults))	-5.69806	0.238162	-23.9251	< 0.0001
					intercept	119.43	0.872521	136.879	< 0.0001
H: Infant mort/1k live births	B: ATMs (per 100k adults)	Middle Africa	0.0018873	5	log(B: ATMs (per 100k adults))	-4.02796	0.674645	-5.97049	0.0018873
					intercept	138.414	0.778272	177.847	< 0.0001

Appendix 2(b) | Back

Trend Lines Model

A linear trend model is computed for average of H: Life exp (years) given average of B: ATMs (per 100k adults). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(B: ATMs (per 100k adults) + intercept)
Number of modeled observations:	31
Number of filtered observations:	24
Model degrees of freedom:	9
Residual degrees of freedom (DF):	22
SSE (sum squared error):	5.47556
MSE (mean squared error):	0.248889
R-Squared:	0.996773
Standard error:	0.498888
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	7	1684.1592	240.594	966.672	< 0.0001

A linear trend model is computed for average of H: Life exp (years) given average of B: Internet users (per 100). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(B: Internet users (per 100) + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45
SSE (sum squared error):	19.9664
MSE (mean squared error):	0.443699
R-Squared:	0.992767
Standard error:	0.666107
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	1787.3967	223.425	503.55	< 0.0001

A linear trend model is computed for average of H: Life exp (years) given average of B: Mobile phone subscribers. The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(B: Mobile phone subscribers + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45
SSE (sum squared error):	23.5988
MSE (mean squared error):	0.524419
R-Squared:	0.991451
Standard error:	0.724168
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	1910.6703	238.834	455.426	< 0.0001

A linear trend model is computed for average of H: Life exp (years) given average of Mobile Usage Density (per 100). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(Avg. Mobile Usage Density (per 100) + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45
SSE (sum squared error):	18.9781
MSE (mean squared error):	0.421736
R-Squared:	0.993125
Standard error:	0.649412
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	2384.0428	298.005	706.615	< 0.0001

A linear trend model is computed for average of H: Life exp (years) given average of B: Scientific & tech articles. The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(B: Scientific & tech articles + intercept)
Number of modeled observations:	50
Number of filtered observations:	5

Strong Correlation

Model degrees of freedom:	10
Residual degrees of freedom (DF):	40
SSE (sum squared error):	25.1947
MSE (mean squared error):	0.629868
R-Squared:	0.989926
Standard error:	0.793642
p-value (significance):	< 0.0001

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	1816.5563	227.07	360.504	< 0.0001

Individual trend lines:

Panes Row	Column	Color Subregion	Line p-value	DF	Coefficients Term	Value	StdErr	t-value	p-value
H: Life exp (years)	B: ATMs (per 100k adults)	Western Africa	N/A	2	B: ATMs (per 100k adults)	0	Residual degrees of freedom for normalized and standard trend line model estimation are different. Full model statistics are unsupported.		
					intercept	55.5833			
H: Life exp (years)	B: ATMs (per 100k adults)	Southern Africa	0.0001666	5	B: ATMs (per 100k adults)	0.157761	0.0156906	10.0545	0.0001666
					intercept	47.9662	0.325158	147.517	< 0.0001
H: Life exp (years)	B: ATMs (per 100k adults)	Northern Africa	0.0774853	5	B: ATMs (per 100k adults)	0.160482	0.072409	2.21633	0.0774853
					intercept	69.3328	0.548898	126.313	< 0.0001
H: Life exp (years)	B: ATMs (per 100k adults)	Middle Africa	0.049949	5	B: ATMs (per 100k adults)	0.138115	0.0537117	2.57142	0.049949
					intercept	51.8785	0.281972	183.984	< 0.0001
H: Life exp (years)	B: ATMs (per 100k adults)	Eastern Africa	0.0015496	5	B: ATMs (per 100k adults)	0.0210966	0.0033817	6.23841	0.0015496
					intercept	54.5609	0.253441	215.28	< 0.0001
H: Life exp (years)	B: Internet users (per 100)	Western Africa	< 0.0001	9	B: Internet users (per 100)	0.562374	0.0704821	7.97896	< 0.0001
					intercept	52.146	0.263948	197.561	< 0.0001
H: Life exp (years)	B: Internet users (per 100)	Southern Africa	0.119169	9	B: Internet users (per 100)	0.293025	0.170164	1.72201	0.119169
					intercept	49.5922	0.837916	59.1852	< 0.0001
H: Life exp (years)	B: Internet users (per 100)	Northern Africa	0.0012291	9	B: Internet users (per 100)	0.0898901	0.0193965	4.63434	0.0012291
					intercept	68.9482	0.267336	257.908	< 0.0001
H: Life exp (years)	B: Internet users (per 100)	Middle Africa	< 0.0001	9	B: Internet users (per 100)	0.492823	0.026076	18.8995	< 0.0001
					intercept	50.4949	0.0858066	588.473	< 0.0001
H: Life exp (years)	B: Internet users (per 100)	Eastern Africa	< 0.0001	9	B: Internet users (per 100)	0.584903	0.0774062	7.55627	< 0.0001
					intercept	51.8068	0.431256	120.13	< 0.0001
H: Life exp (years)	B: Mobile phone subscribers	Western Africa	< 0.0001	9	B: Mobile phone subscribers	3.878e-007	4.47e-008	8.67579	< 0.0001
					intercept	52.3941	0.222827	235.134	< 0.0001
H: Life exp (years)	B: Mobile phone subscribers	Southern Africa	0.166089	9	B: Mobile phone subscribers	1.351e-007	8.968e-008	1.50696	0.166089
					intercept	50.0809	0.647226	77.3777	< 0.0001
H: Life exp (years)	B: Mobile phone subscribers	Northern Africa	0.00156	9	B: Mobile phone subscribers	8.263e-008	1.849e-008	4.46762	0.00156
					intercept	68.9768	0.271172	254.365	< 0.0001
H: Life exp (years)	B: Mobile phone subscribers	Middle Africa	< 0.0001	9	B: Mobile phone subscribers	5.811e-007	1.95e-008	29.8057	< 0.0001
					intercept	50.9806	0.0415288	1227.6	< 0.0001
H: Life exp (years)	B: Mobile phone subscribers	Eastern Africa	0.0003277	9	B: Mobile phone subscribers	8.605e-007	1.532e-007	5.61559	0.0003277
					intercept	52.9571	0.411931	128.558	< 0.0001
H: Life exp (years)	Mobile Usage Density (per 100)	Western Africa	< 0.0001	9	Avg. Mobile Usage Density (per 100)	0.0742149	0.0088377	8.39755	< 0.0001
					intercept	52.326	0.23556	222.134	< 0.0001
H: Life exp (years)	Mobile Usage Density (per 100)	Southern Africa	0.0541665	9	Avg. Mobile Usage Density (per 100)	0.0254033	0.0114784	2.21315	0.0541665
					intercept	50.0254	0.495587	100.942	< 0.0001
H: Life exp (years)	Mobile Usage Density (per 100)	Northern Africa	0.0013413	9	Avg. Mobile Usage Density (per 100)	0.023203	0.005074	4.57293	0.0013413
					intercept	68.9458	0.270882	254.523	< 0.0001
H: Life exp (years)	Mobile Usage Density (per 100)	Middle Africa	< 0.0001	9	Avg. Mobile Usage Density (per 100)	0.0524014	0.002299	22.793	< 0.0001
					intercept	50.8814	0.0574569	885.557	< 0.0001
H: Life exp (years)	Mobile Usage Density (per 100)	Eastern Africa	< 0.0001	9	Avg. Mobile Usage Density (per 100)	0.15296	0.021147	7.23316	< 0.0001
					intercept	52.1044	0.415789	125.315	< 0.0001
H: Life exp (years)	B: Scientific & tech articles	Western Africa	0.0028593	8	B: Scientific & tech articles	0.248178	0.0586103	4.23438	0.0028593
					intercept	41.8931	2.75779	15.1908	< 0.0001
H: Life exp (years)	B: Scientific & tech articles	Southern Africa	0.152633	8	B: Scientific & tech articles	0.0086489	0.005472	1.58056	0.152633
					intercept	46.3346	2.84001	16.3149	< 0.0001
H: Life exp (years)	B: Scientific & tech articles	Northern Africa	0.008201	8	B: Scientific & tech articles	0.0067498	0.0019342	3.48978	0.008201
					intercept	66.1489	1.04317	63.4114	< 0.0001
H: Life exp (years)	B: Scientific & tech articles	Middle Africa	0.0039291	8	B: Scientific & tech articles	0.145579	0.0363606	4.00376	0.0039291
					intercept	48.8203	0.740027	65.971	< 0.0001
H: Life exp (years)	B: Scientific & tech articles	Eastern Africa	0.0015266	8	B: Scientific & tech articles	0.225667	0.0479375	4.70752	0.0015266
					intercept	43.2546	2.35867	18.3386	< 0.0001

Appendix 3(a) | Back

Trend Lines Model

A logarithmic trend model is computed for average of H: Infant mort/1k live births given average of F: GDP per capita (curr $). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(log(F: GDP per capita (curr $)) + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45
SSE (sum squared error):	669.712
MSE (mean squared error):	14.8825
R-Squared:	0.990798
Standard error:	3.85778
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	45639.105	5704.89	383.329	< 0.0001

A logarithmic trend model is computed for average of H: Infant mort/1k live births given average of H: Health exp/cap (curr $). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(log(H: Health exp/cap (curr $)) + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45
SSE (sum squared error):	547.205
MSE (mean squared error):	12.1601
R-Squared:	0.992481
Standard error:	3.48713
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	38412.316	4801.54	394.86	< 0.0001

An exponential trend model is computed for natural log of average of H: Infant mort/1k live births given sum of International Tourism Profit. The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula:	Subregion*(International Tourism Profit + intercept)
Number of modeled observations:	55
Number of filtered observations:	0
Model degrees of freedom:	10
Residual degrees of freedom (DF):	45
SSE (sum squared error):	0.338167
MSE (mean squared error):	0.0075148
R-Squared:	0.966594
Standard error:	0.0866881
p-value (significance):	< 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	2.2591556	0.282394	37.5783	< 0.0001

Individual trend lines:

Panes		**Color**	**Line**		**Coefficients**				
Row	**Column**	**Subregion**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
H: Infant mort/1k live births	H: Health exp/cap (curr $)	Western Africa	< 0.0001	9	log(H: Health exp/cap (curr $))	-29.8025	2.05828	-14.4793	< 0.0001
					intercept	240.544	7.16143	33.5889	< 0.0001
H: Infant mort/1k live births	H: Health exp/cap (curr $)	Southern Africa	0.0002276	9	log(H: Health exp/cap (curr $))	-29.7347	5.03525	-5.90529	0.0002276
					intercept	245.33	27.3302	8.97651	< 0.0001
H: Infant mort/1k live births	H: Health exp/cap (curr $)	Northern Africa	< 0.0001	9	log(H: Health exp/cap (curr $))	-13.475	1.55813	-8.64821	< 0.0001
					intercept	110.763	7.59057	14.5922	< 0.0001
H: Infant mort/1k live births	H: Health exp/cap (curr $)	Middle Africa	< 0.0001	9	log(H: Health exp/cap (curr $))	-9.75571	0.463673	-21.0401	< 0.0001
					intercept	181.38	2.04263	88.7973	< 0.0001
H: Infant mort/1k live births	H: Health exp/cap (curr $)	Eastern Africa	< 0.0001	9	log(H: Health exp/cap (curr $))	-52.9893	3.72534	-14.224	< 0.0001
					intercept	327.015	15.5284	21.0592	< 0.0001
H: Infant mort/1k live births	F: GDP per capita (curr $)	Western Africa	< 0.0001	9	log(F: GDP per capita (curr $))	-33.7576	2.20826	-15.2869	< 0.0001
					intercept	353.04	14.1235	24.9967	< 0.0001
H: Infant mort/1k live births	F: GDP per capita (curr $)	Southern Africa	0.0003365	9	log(F: GDP per capita (curr $))	-36.5136	6.52639	-5.59476	0.0003365
					intercept	377.644	52.4599	7.1987	< 0.0001
H: Infant mort/1k live births	F: GDP per capita (curr $)	Northern Africa	0.0016551	9	log(F: GDP per capita (curr $))	-12.3021	2.77907	-4.42669	0.0016551
					intercept	143.416	22.1888	6.46346	0.0001163
H: Infant mort/1k live births	F: GDP per capita (curr $)	Middle Africa	< 0.0001	9	log(F: GDP per capita (curr $))	-9.41838	1.04987	-8.971	< 0.0001
					intercept	212.325	8.22169	25.825	< 0.0001
H: Infant mort/1k live births	F: GDP per capita (curr $)	Eastern Africa	< 0.0001	9	log(F: GDP per capita (curr $))	-50.8891	3.99331	-12.7436	< 0.0001
					intercept	471	28.6233	16.4552	< 0.0001
H: Infant mort/1k live births	International Tourism Profit	Western Africa	0.0021352	9	International Tourism Profit	2.011e-011	4.729e-012	4.25229	0.0021352
					intercept	4.9836	0.0213528	233.393	< 0.0001
H: Infant mort/1k live births	International Tourism Profit	Southern Africa	0.445958	9	International Tourism Profit	-3.273e-011	4.106e-011	-0.797016	0.445958
					intercept	4.52031	0.134797	33.5342	< 0.0001
H: Infant mort/1k live births	International Tourism Profit	Northern Africa	< 0.0001	9	International Tourism Profit	-2.426e-011	2.073e-012	-11.7019	< 0.0001
					intercept	4.07889	0.024795	164.504	< 0.0001
H: Infant mort/1k live births	International Tourism Profit	Middle Africa	0.207878	9	International Tourism Profit	-5.287e-011	3.897e-011	-1.35686	0.207878

					intercept	4.91145	0.0194133	252.995	< 0.0001
H: Infant mort/1k live births	International Tourism Profit	Eastern Africa	< 0.0001	9	International Tourism Profit	-6.472e-011	5.088e-012	-12.7218	< 0.0001
					intercept	4.86606	0.0177824	273.644	< 0.0001

Appendix 3(b) | Back

Trend Lines Model

An exponential trend model is computed for natural log of average of H: Life exp (years) given average of International Tourism Profit. The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula: Subregion*(Avg. International Tourism Profit + intercept)
Number of modeled observations: 55
Number of filtered observations: 0
Model degrees of freedom: 10
Residual degrees of freedom (DF): 45
SSE (sum squared error): 0.0113164
MSE (mean squared error): 0.0002515
R-Squared: 0.985114
Standard error: 0.015858
p-value (significance): < 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	0.30460243	0.0380753	151.408	< 0.0001

A logarithmic trend model is computed for average of H: Life exp (years) given average of F: GDP per capita (curr $). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula: Subregion*(log(F: GDP per capita (curr $)) + intercept)
Number of modeled observations: 55
Number of filtered observations: 0
Model degrees of freedom: 10
Residual degrees of freedom (DF): 45
SSE (sum squared error): 22.6605
MSE (mean squared error): 0.503567
R-Squared: 0.991791
Standard error: 0.709624
p-value (significance): < 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	2519.2901	314.911	625.361	< 0.0001

A logarithmic trend model is computed for average of H: Life exp (years) given average of H: Health exp/cap (curr $). The model may be significant at p <= 0.05. The factor Subregion may be significant at p=0.05.

Model formula: Subregion*(log(H: Health exp/cap (curr $)) + intercept)
Number of modeled observations: 55
Number of filtered observations: 0
Model degrees of freedom: 10
Residual degrees of freedom (DF): 45
SSE (sum squared error): 18.5678
MSE (mean squared error): 0.412618
R-Squared: 0.993273
Standard error: 0.642354
p-value (significance): < 0.0001

Strong Correlation

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Subregion	8	2627.4219	328.428	795.96	< 0.0001

Individual trend lines:

Panes		Color	Line		Coefficients				
Row	Column	Subregion	p-value	DF	Term	Value	StdErr	t-value	p-value
H: Life exp (years)	F: GDP per capita (curr $)	Western Africa	< 0.0001	9	log(F: GDP per capita (curr $))	4.53434	0.296915	15.2715	< 0.0001
					intercept	24.7849	1.89899	13.0516	< 0.0001
H: Life exp (years)	F: GDP per capita (curr $)	Southern Africa	0.380329	9	log(F: GDP per capita (curr $))	0.911862	0.988459	0.922509	0.380329
					intercept	43.6218	7.94536	5.49022	0.000385
H: Life exp (years)	F: GDP per capita (curr $)	Northern Africa	0.0678714	9	log(F: GDP per capita (curr $))	1.70084	0.819895	2.07446	0.0678714
					intercept	56.3206	6.54625	8.6035	< 0.0001
H: Life exp (years)	F: GDP per capita (curr $)	Middle Africa	0.0003498	9	log(F: GDP per capita (curr $))	1.34408	0.24154	5.5646	0.0003498
					intercept	41.3801	1.89154	21.8764	< 0.0001
H: Life exp (years)	F: GDP per capita (curr $)	Eastern Africa	< 0.0001	9	log(F: GDP per capita (curr $))	9.06911	0.849053	10.6814	< 0.0001
					intercept	-10.4448	6.08584	-1.71625	0.12025
H: Life exp (years)	H: Health exp/cap (curr $)	Western Africa	< 0.0001	9	log(H: Health exp/cap (curr $))	4.00595	0.272089	14.7229	< 0.0001
					intercept	39.8854	0.946687	42.1316	< 0.0001
H: Life exp (years)	H: Health exp/cap (curr $)	Southern Africa	0.367318	9	log(H: Health exp/cap (curr $))	0.753348	0.793682	0.949182	0.367318
					intercept	46.8677	4.30793	10.8794	< 0.0001
H: Life exp (years)	H: Health exp/cap (curr $)	Northern Africa	0.0090101	9	log(H: Health exp/cap (curr $))	2.12807	0.641929	3.31512	0.0090101
					intercept	59.5469	3.12723	19.0415	< 0.0001
H: Life exp (years)	H: Health exp/cap (curr $)	Middle Africa	< 0.0001	9	log(H: Health exp/cap (curr $))	1.46877	0.124035	11.8416	< 0.0001
					intercept	45.4618	0.546414	83.2003	< 0.0001
H: Life exp (years)	H: Health exp/cap (curr $)	Eastern Africa	< 0.0001	9	log(H: Health exp/cap (curr $))	9.45475	0.804358	11.7544	< 0.0001
					intercept	15.1679	3.35282	4.52393	0.0014387
H: Life exp (years)	International Tourism Profit	Western Africa	0.0018313	9	Avg. International Tourism Profit	-8.322e-011	1.91e-011	-4.35707	0.0018313
					intercept	3.96318	0.0070068	565.622	< 0.0001
H: Life exp (years)	International Tourism Profit	Southern Africa	0.247646	9	Avg. International Tourism Profit	-2.855e-011	2.309e-011	-1.23628	0.247646
					intercept	3.9481	0.0154317	255.844	< 0.0001

H: Life exp (years)	International Tourism Profit	Northern Africa	0.0035777	9	Avg. International Tourism Profit	1.384e-011	3.542e-012	3.90761	0.0035777
					intercept	4.22027	0.0073418	574.83	< 0.0001
H: Life exp (years)	International Tourism Profit	Middle Africa	0.338786	9	Avg. International Tourism Profit	6.517e-011	6.451e-011	1.0102	0.338786
					intercept	3.95319	0.0067814	582.945	< 0.0001
H: Life exp (years)	International Tourism Profit	Eastern Africa	< 0.0001	9	Avg. International Tourism Profit	2.946e-010	3.606e-011	8.16819	< 0.0001
					intercept	3.92859	0.0094417	416.092	< 0.0001