PYKNOT – Going The Distance

SUMMARY REPORTS
[SEPT 2019]

TABLE OF CONTENT

Olympic and Major Marathon Performance

Since our overall interest was in establishing the dominance of African countries in the marathon, the first task was accessing and cleaning all of our data. For our project, we used datasets from Kaggle that provided 100 Years of Olympic Data and the World Marathon Majors. The process of cleaning the data was more complicated for the world marathon data since it was provided in separate CSV files for men and women per each marathon. Most of the files were consistent with the same information but there was some standardization involved before concatenating. With the Olympic data, we had to specify which events to include and select the countries we were interested in. There was a phenomenon that became apparent in the trajectory of wins for each of the countries. We knew about the tribe of Kenyan runners who were dominating the major distance running events. However, seeing the shift from Western nations to African countries begged the question, why?.

The US has been a top performer since the start of the modern games earning half of their Olympic medals before any African nations had placed. In spite of having the largest medal count and competitive athletes, Ethiopia—who emerged as a top competitor in the 60's—and Kenya—who has accumulated the second most medals since the late 80's—are seemingly unparalleled in recent distance running events. The Olympic data did not provide times for all of these medalists, which would have been interesting to compare with the marathons of the Abbot World Majors. The marathon data from the Majors showed that Kenyan and Ethiopian runners were top tier in comparison to other countries. Not only do they hold the top times for each major race, but they are consistently running at record-shattering times. Overall, this analysis helped us frame the rest of our questions. The data demonstrated that these African nations are dominant in the marathon. What were the factors that could explain their success?

GDP VS Olympic performance:

The purpose was to find a relationship between the country's GDP and the number of medals they earned. We attempted to discover with the data if certain African Countries skewed to winning more medals respective to their GDP. This experiment was interesting primarily from the data gathering procedure because the resources used were from unique sources. This

resulted in data not forming together seamlessly like it would from the homework. Two things were apparent. One: there was a clear relationship between the number of medals a country won and the country's GDP. Furthermore, we were able to see that the african countries skewed to receive more Medals with respect to their GDP. Although this was a small change, and might fail to be proven statistically significant, there did seem to be a real relationship.

GDP Vs Marathon Wins:

The starkest difference between the US, UK, Germany, Japan, Kenya and Ethiopia is their economies. After plotting the bar graph for the top country marathon wins in the last 5 decades, it was amazing to see that the dominance of African nations started in the 1980s. We decided to explore GDP growth rates amongst the top 5 country marathon winners which are Kenya, the United States, Ethiopia, the United Kingdom and Japan looking for answers.

First off, all these countries had an averagely increasing GDP growth rate on a nominal basis. Which meant that national incomes were not rebased year in year out and inflation/deflation was not accounted for. This was quite a considerable limitation to the analysis. Another challenge we encountered in cleaning the data was determining when African nations started participating in the major marathons and the number of runners allowed to represent a country. Despite the challenges, we still proceeded with the analysis using a scatter plot. This showed a distinct negative correlation between the total number of African nation-wins (Kenya & Ethiopia) and the developed nation-wins (USA, UK & Japan) vis-à-vis their growing GDP in the last 5-6 decades. We relied on the scatter plot for visualizing our data because it gives a rather optimum representation of the wins vis-à-vis the galloping differences in GDP between the developed countries in question and African nations.

Weather, environment and demographics:

We were interested in exploring how the visualization of data could bring context to the sudden rise and domination of endurance athletes from the Great Rift Valley. Both Kenyan and Ethiopian produce endurance athletes from cities located in this geographic feature which cuts through both countries and beyond. The demographic data provided by the CIA World Factbook is expansive, providing ample resources for understanding the makeup of this region. While there was not an opportunity to use an API, there is a considerable Json dump on Github. In order to prevent scope creep focus remained on economic and geographic data. We considered data on Infrastructure, health and even text descriptions involving geography and climate. In addition to data we searched for publications on topics inline with our query. What was found indicated a robust culture of running pervasive in the area. There were also lines of inquiry into how poverty, lack of developed infrastructure, climate and geography all acted as determining factors in the performance of athletes from the Great Rift Valley. What was not discovered was much, if any, consideration of data in the observations these authors shared.

Using an API call to gather historical weather data we explored the climate and conditions of the races, Ethiopia and Kenya. There were challenges in making pertinent and reliable assessments about the influence of weather. Heat index would seemed to be the best comparison of the conditions the athletes faced during the Abbott World Marathon Majors and their training regimens. The time frame considered—2013 to 2018—presented the most consistent set of data shared by all the marathon events and the nations we focused on. Training for marathons is a continuous endeavor and the competitions in question last a matter of hours. In order to account for weather conditions while training in the Great Rift Valley, the selected dates corresponded with the solstices and equinoxes for the capital cities of the nations from Africa. We also examined the elevations of the host cities of the Majors and of Iten in Kenya as well as Bekoji in Ethiopia.

The daily grind of life requires many citizens of Kenya and Ethiopia to travel by foot. As a result, children in these countries tend to run to school. Many of the publications assumed this explained the larger pool of elite distance runners from these countries. Yet, rarely was there any evidence provided. Paved roads facilitate travel by automobiles and CIA WFB tracked this data and ranked countries by this metric. This seemed to corroborate the supposition that athletes from this area were prone to running as a necessity. Data on improved water sources—indoor plumbing—also confirmed this idea. Humans, needing water to live, will either live near it or be forced to gather it. Since this region lacked such infrastructure as compared to the other countries we considered it stands to reason that considerable time would be spent walking in order to secure the necessities of life. Urbanization was another metric used since infrastructure density is greatest in cities. Most of Kenya and Ethiopia consisted of rural areas exacorgabting the two aforementioned challenges.

In regards to climate the data didn't suggest any significant difference between the weather data for the region when compared to the conditions during the race events we examined. There is more data available than what was examined that could lead to some insights. More time would be needed to explore that. As far as elevation was concerned, there was considerable differences between the region we study and the cities where these events were held. This suggests an advantage for the athletes from this region. However, we did not consider elevations of the regions the rest of the competitors hailed from, only the cities that held the races. Much more detail would have to be gathered to make any conclusions here.

There are strong arguments to be made for the influence the lack of development could have on the performance on endurance athletes from Kenya and Ethiopia. Many considerations —desire for social mobility, celebrity culture, societal pressures—should be explored before anything conclusive on the benefits of poverty and poor economic development could posited. However, the data we gathered is compelling. There is something about the athletes from Ethiopia and Kenya that makes one pause. There accomplishments are too pronounced and the disparities between these countries and the rest of the developed world to not wrestled with.
What makes these athletes from the Great Rift Valley keep going the distance?