



ARRIA + TIBCO WHITE PAPER

TIBCO SPOTFIRE DASHBOARD AUGMENTED WITH ARRIA NLG

ARRIA

TIBCO SPOTFIRE DASHBOARD AUGMENTED WITH ARRIA NLG

Introduction

IN THIS DIGITAL AGE, electronic data is a more important source of information than our own senses to understand the happenings around us.

For example, our senses have limited use in the context of the COVID-19 pandemic, and we seek timely and accurate data to understand the behavior of the outbreak because each of us realizes the importance of this data for our own individual survival. However, raw data on its own is no good because we lack either the necessary data skills or the necessary time to manually inspect every bit of data. During these extraordinary times we now live in, each one of us seeks technological help. From the robust roots of data models and query languages in the 1970s, data related technologies have come a long way.

In 2020, we now have technologies such as data analytics, interactive dashboards, and natural language generation (NLG) that can discover and tell data stories automatically. Because data stories show us the human face of data aligning well with our mental models of the world, these new technologies manage to hide the much-dreaded number crunching away from us. Very much like traditional stories, well-told data stories are made up of a balanced number of visuals and linguistic descriptions. The linguistic descriptions serve to tie the visuals together into a coherent data story.

In this white paper, we describe the technical details of the COVID-19 Live Dashboard which brings together Arria's NLG technology and TIBCO's Visual Analytics technology.



Contents

Introduction	2
What is NLG?	3
What are Visual Analytics?	3
NLG + Visual Analytics.....	3
Real-World Use Case	4
Global Status Check.....	5
Latest Growth Trajectory	7
Latest R0 Estimates	8

TIBCO SPOTFIRE DASHBOARD AUGMENTED WITH ARRIA NLG

A very important outcome of creating linguistic data stories is that they can be directly interpreted.

What is NLG?

Language is the most sophisticated medium for information transfer among humans. Natural Language Generation (NLG) is an innovative AI technology that translates data into linguistic data stories. Arria Studio is a platform for programming linguistic data stories. An NLG program developed in Arria Studio mainly specifies mappings from data (or data-derived abstractions) to linguistic constructs such as words, phrases and clauses. A very important outcome of creating linguistic data stories is that they can be directly interpreted. This means, linguistic data stories truly democratize the data riches to every one of us.

What are Visual Analytics?

Because we, humans, have a powerful visual perception system, visual representations of data offer an effective medium to communicate large volumes of information. In addition, interactions with data can be effectively modeled as interactions with their corresponding visuals. Therefore, interactive visuals emerged as the human face of data – allowing users to interact with data visually, query data visually, and view results of these queries visually as well. In addition, sophisticated data analytics can be invoked to build models of data displayed visually—through charts, maps, graphs and color-coding—offering rich data exploration power to users. However, not everyone is skilled enough or has the necessary time to exploit all these powerful data exploration capabilities.

NLG + Visual Analytics

A major strength of Arria's technology is that NLG applications in Arria NLG Studio can be integrated into a wide range of data analytics platforms such as TIBCO Spotfire, a Visual Analytics platform. This integration offers interactive linguistic data stories that are accessible to all users irrespective of their data skills, while allowing skilled users to explore data exhaustively using visual analytics.

TIBCO SPOTFIRE DASHBOARD AUGMENTED WITH ARRIA NLG

By combining visuals and narrative, a range of users can quickly understand the information they are looking for.

Real World Use Case

The COVID-19 live report is a web-based dashboard that uses Visual Analytics and NLG to give people up-to-date information about the COVID-19 virus. It is intended to be a resource for anyone who is interested in learning more about the state of the pandemic.

Following, we describe some of the sections of the dashboard, and how they integrate NLG with Visual Analytics.



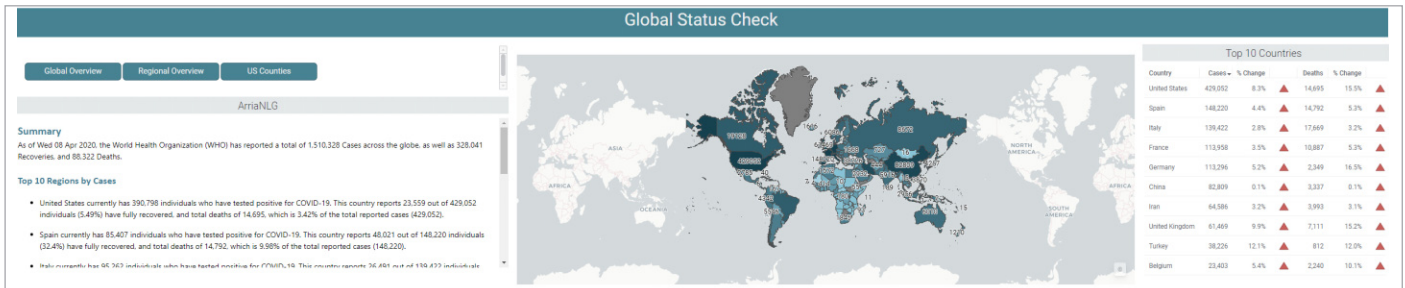
Global Status Check

The Global Status Check shows the spread of COVID-19 throughout the world. It consists of a **map** which is color-coded to show roughly how many cases are in each country, a **table** which lists the ten countries with the most cases, and a **narrative** produced by Arria NLG which gives more insights into the top ten countries. The narrative, map, and table serve different purposes as noted on the following page.

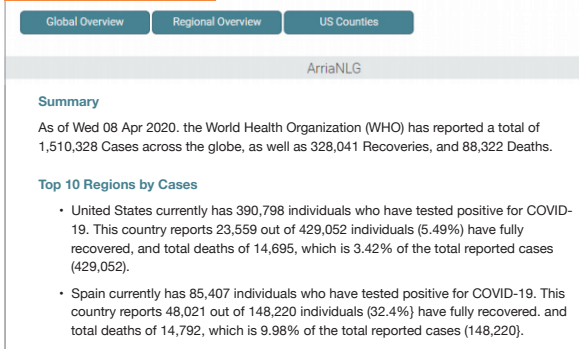
In short, the combination of the BI presentations (map and table) and the narrative is very powerful, and lets a wide variety of users quickly understand the information they are looking for.

See next page for more details.

Global Status Check Detail



THE NARRATIVE



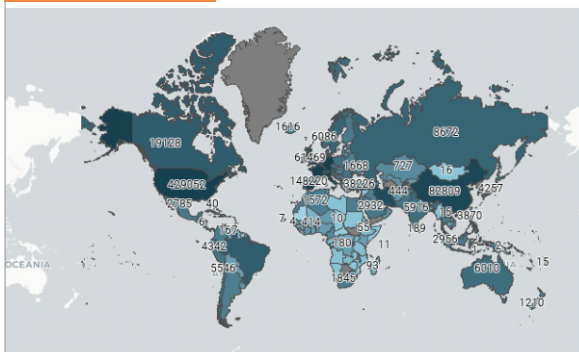
This Dashboard panel is comprised of three parts that serve different purposes to show the spread of COVID-19 throughout the world.

1) The Narrative gives more detailed information about selected countries; it's less useful for comparisons than the map or table, but it does a better job of describing what is happening with one country. It's also easier for "verbal" (as opposed to "visual") thinkers to understand.

2) The Map lets users quickly spot regions that have a lot of reported cases of coronavirus, and regions that have so far been spared the worst ravages of the virus.

3) The Table lets users quickly get more information about selected countries. It's especially useful for comparisons, for example, we can quickly see the differences between the United States and China.

THE MAP



THE TABLE

Top 10 Countries				
Country	Cases	% Change	Deaths	% Change
United States	429,052	8.3%	14,695	15.5%
Spain	148,220	4.4%	14,792	5.3%
Italy	139,422	2.8%	17,669	3.2%
France	113,958	3.5%	10,887	5.3%
Germany	113,296	5.2%	2,349	16.5%
China	82,809	0.1%	3,337	0.1%
Iran	64,596	3.2%	3,993	3.1%

TIBCO SPOTFIRE DASHBOARD AUGMENTED WITH ARRIA NLG

Having both visual and narrative means people with different backgrounds can immediately benefit.

Latest Growth Trajectory

Latest Growth Trajectory shows how COVID-19 is spreading and growing across the world. It consists of a line graph showing case growth, a narrative summarizing what is happening in each continent, as well as some buttons for exploring more details. Again, the combination of the line graph and narrative is powerful, and lets a variety of different users quickly find key information.

As with the Global Status Check, having both presentations means that people of different backgrounds can benefit from the dashboard, including both “visual” thinkers who might find the narrative to be wordy, and “verbal” thinkers who might find the line graph a bit overwhelming because it is simultaneously showing data from many different countries. Users can click on the buttons to get more details about specific countries, states, and provinces

See next page for more details.

Latest R0 Estimates

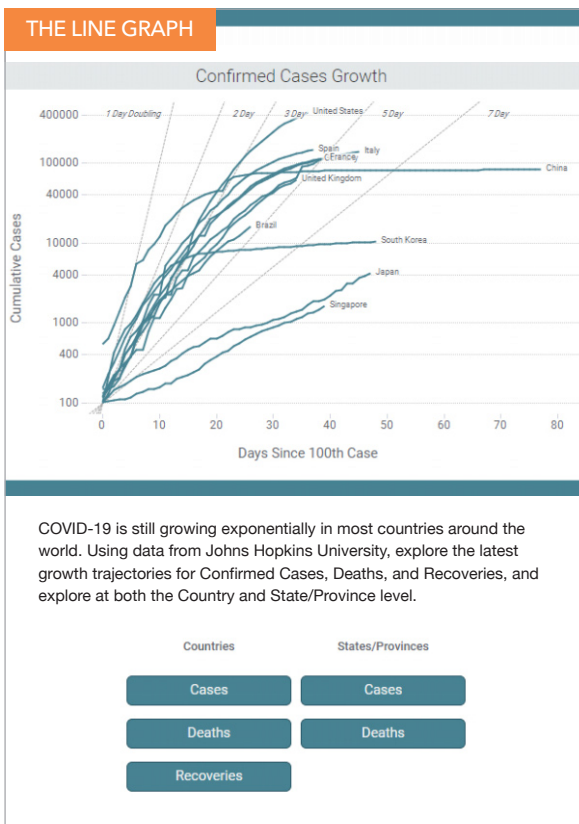
The Latest R0 Estimates dashboard panel is explained on page 8. R0 is a key indicator of how fast COVID-19 is spreading and/or declining. Again, the COVID-19 Live Dashboard shows the power of combining a visual presentation **map** with a **narrative**.

Users can click on the buttons to get more details about specific countries, states, and provinces.

This results in a line graph (similar to the above), a bar chart, and a detailed narrative about a specific country.

See page 8 for more details.

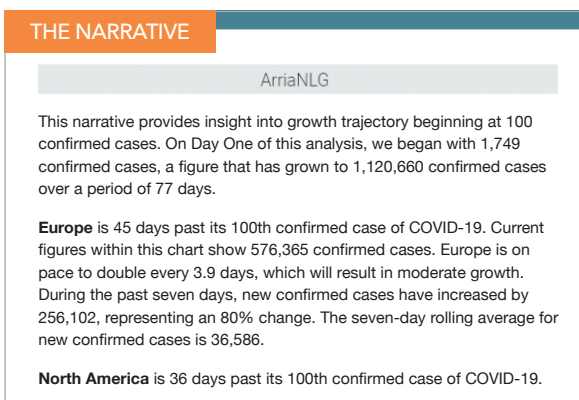
Latest Growth Trajectory



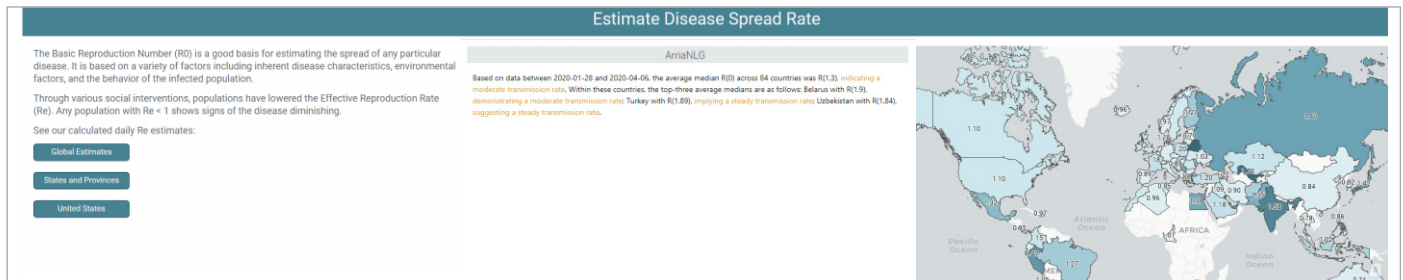
This Dashboard panel is comprised of two parts that show how COVID-19 is both spreading and growing across the world.

1) The Line Graph is an excellent tool for understanding the progression of the virus in different countries.

2) The Narrative gives detailed information and insights about the progression of the virus within each continent.



Latest R0 Estimates



THE NARRATIVE

The Basic Reproduction Number (R_0) is a good basis for estimating the spread of any particular disease. It is based on a variety of factors including inherent disease characteristics, environmental factors, and the behavior of the infected population.

Through various social interventions, populations have lowered the Effective Reproduction Rate (R_e). Any population with $R_e < 1$ shows signs of the disease diminishing.

See our calculated daily R_e estimates:

Global Estimates

States and Provinces

United States

ArriaNLG

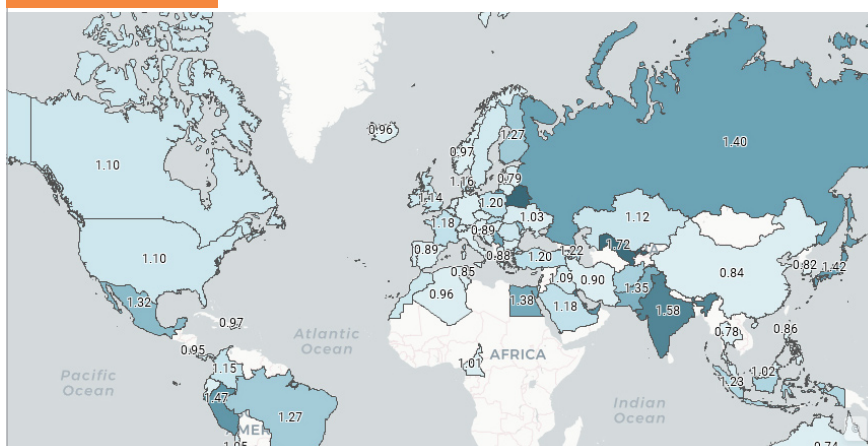
Based on data between 2020-01-28 and 2020-04-06, the average median R_0 across 84 countries was $R(1.3)$, indicating a moderate transmission rate. Within these countries, the top-three average medians are as follows: Belarus with $R(1.9)$, demonstrating a moderate transmission rate; Turkey with $R(1.89)$, implying a steady transmission rate; Uzbekistan with $R(1.84)$, suggesting a steady transmission rate.

This Dashboard panel is comprised of two parts:

1) The Narrative gives more detailed information about what is happening worldwide and in the countries with the highest R_0 .

2) The Map makes it easy to see which regions of the world have the highest R_0 . As always with maps, this representation of data is better at showing the situation in a region than in a specific country, especially if the country is small.

THE MAP



Arria NLG integration with TIBCO dashboards offers interactive linguistic data stories that are accessible to all users irrespective of their data skills, while allowing skilled users to explore data exhaustively using visual analytics.

For more information, including a demo tailored to your specific use case, please email: sales@arria.com
arria.com

UNITED STATES | CANADA | UNITED KINGDOM | EMEA | AUSTRALIA | NEW ZEALAND

ABOUT THE AUTHORS



Prof. Ehud Reiter, Arria Chief Scientist

Prof. Reiter is a pioneer in the science of Natural Language Generation (NLG) and one of the world's foremost authorities in the field of NLG. He is responsible for the overall direction of Arria's core technology development as well as supervision of specific NLG projects. He is Professor of Computing Science in the University of Aberdeen School of Natural and Computing Sciences.



Dr. Yaji Sripada, Arria Chief Scientist

Dr. Sripada is recognized globally as a leading NLG expert. He is widely published on the subject of NLG and his papers are frequently cited. He is responsible for Arria's engineering and software applications and supervision of specific NLG projects. Dr. Sripada specializes in integrating NLG with adjacent technologies, such as data analytics and information visualization. He is also a Senior Lecturer in Computing Science at the University of Aberdeen.