SAS® GLOBAL FORUM 2020 #SASGF

Paper SAS4126-2020

Best Practices for Effective Infographics in SAS® Visual Analytics

Falko Schulz, SAS Institute Inc.

ABSTRACT

SAS® Visual Analytics enables you to build impressive infographics into your report. Learn how to effectively convey your message using the out-of-the box visualizations and layouts delivered with SAS Visual Analytics. In this paper, the author will analyze Earth's surface temperatures using daily and monthly NASA/GISS data for over 25,000 stations across the globe. We will learn how temperatures changed over the years and detect trends and hotspots. You will learn how to import and prepare data in SAS Viya, as well as discover the latest visualization capabilities. We will discuss best practices for designing and developing simple but effective data visualizations targeted for specific audiences.

INTRODUCTION

Infographics deliver information with clarity and simplicity. With the growing amount of data available and expectations to communicate insights from data better, it is important to understand the basic concepts of visual design and key considerations when building reports targeted for specific audiences. This paper covers tips and techniques to master effective report design using SAS Visual Analytics to develop and design high-impact data visualizations.

DASHBOARD OR INFOGRAPHIC

A dashboard is described as "a user interface that, somewhat resembling an automobile's dashboard, organizes and presents information in a way that is easy to read. However, a computer dashboard is more likely to be interactive than an automobile dashboard." (WhatIs.com 2018). The focus in dashboards is about providing a comprehensive view of one or more data perspectives. Often dashboards provide additional user-driven context and real-time interactivity during exploration. A modern version of the dashboard behaves more like an information application and provides a rich and complete data experience.

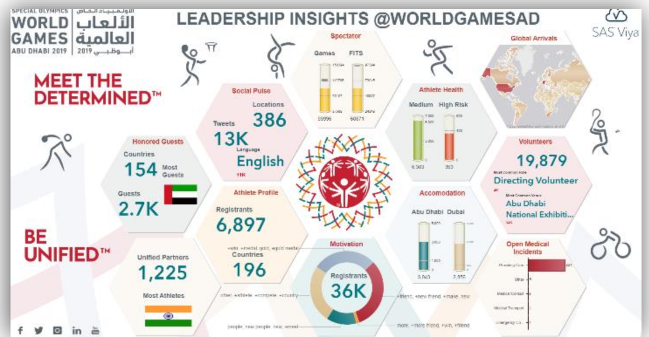
An infographic on the other hand is a "graphic visual representations of information, data, or knowledge intended to present information quickly and clearly" [Wikipedia 2018]. The focus here lies on a balanced weight between 'info' (textural information) and 'graphic' (graphical) representation of data. Often designers use a combination of data and art to best present numbers in form of shapes or graphics. It's the resulting visual impact that differentiates compelling from less engaging infographics.

AUDIENCES

While a great dashboard is designed once and used by many users in many contexts, infographics typically target specific audiences to provide a custom experience dependent on user knowledge and status. For example, you might have dedicated reports for general public versus management by providing different levels of information and aggregation.

BEST PRACTICES

Creating infographics is not difficult given some basic creativity and some key points to consider when doing so:



Identify audience – Be specific. For example, general public versus management.

Simplicity –Minimize number of visual types and use the most simplified representation (for example, KPI versus list table). If you have content for multiple stories, then chances are you need more than one infographic.

Focus – Use techniques to draw attention to the key areas. Stay on one topic or story.

Headline – Design it as big and bold statement. Use a short but descriptive banner text or graphic.

Balance – Use both 'info' and 'graphics' to tell your story with the weight on the visuals.

Art – Get creative. Use white space. Use story-matching 'simple' color themes. Decide on a consistent font style.

Flow – Readers start at the headline then work their way through all of the linked visuals.

Double check – Obvious one. Double-check all of the facts and sources.

Cite – Be sure to cite sources and include a footer with author details.

THE TOOL

The paper focuses on the SAS flagship report and design product, SAS Visual Analytics, which provides an excellent set of data visualization and design capabilities. While you might not need all available features, the tool provides a unified user experience in all stages of design - from data preparation, visualization exploration, and design to collaborative information sharing.

SAS Visual Analytics provides almost unlimited ways to explore and visualize data. It's used by many organizations for big data exploration, as well as self-service analytics and dashboards. Some key features of the SAS Viya releases include:

- Location analytics – Brings your data to life with demographic enrichment, automatic clusters, custom polygons, and more.
- Custom graphs – Extends your content by using third-party graphs as if they were native graphs. Graph libraries like D3 or C3 are now able to be used in SAS Visual Analytics.
- Ambient analytics –Includes so much self-service analysis that it delivers ambient analytics, which allows business teams to grow their skills and create dashboards that are powered by advanced analytics.
- More Visuals – Enables you to tell your story using out-of-the-box visuals. There are many more visualizations to use to represent the data just the way you want.

- Open and extensible – Provides new APIs to enable you to call the elements of the dashboard and share it on the https://developer.sas.com site. This means you can create endless entry points into SAS Visual Analytics.

Here are some general tips for building your infographic-styled dashboards using SAS Visual Analytics:

- Drop zones and containers – You can achieve a whole lot of formatting by just dragging and dropping in SAS Visual Analytics today using intelligent drop zones without the need for containers. You can use the container object in SAS Visual Analytics to group objects together. There are benefits in using both drop zones and container objects. However, when doing precise designs, one or more container objects will provide many options to format and control layout.
- Object overlay – A key feature to take advantage of in SAS Visual Analytics when designing infographic dashboards is the use of object overlay in precision containers. The container object facilitates this concept. However, the design effect it allows is impressive. This paper covers various examples, where the author places images over graphs, or a graph on another graph. This allows for very creative designs with SAS Visual Analytics and can be the difference between a good dashboard and a great dashboard.
- Interactivity versus static – Understanding how your infographic will be used is going to improve your build process. Typical dashboards can often be the only tool a user must use to slice and dice data, so interactivity might need to be embedded in the design. If the infographic is built only for a single view of the data, then a static view will suffice. However, SAS Visual Analytics allows flexible interactive navigation for the user if it's enabled using the related options in the Actions pane.
- One report, many users – Understanding if you are designing an infographic for more than one user group can affect the design and might require some additional features to be used. For example, data-driven content objects or parameters assisted navigation allows flexible rendering at run time.
- New graphs to tell better stories – SAS Visual Analytics has many more out-of-the-box visualizations to use in your design like the *key value* object, which has an infographic setting to create quick and simple infographic styled elements in your dashboard. You may also leverage many more great visualizations made by others around the globe, given the flexibility the *data-driven content* object provides.

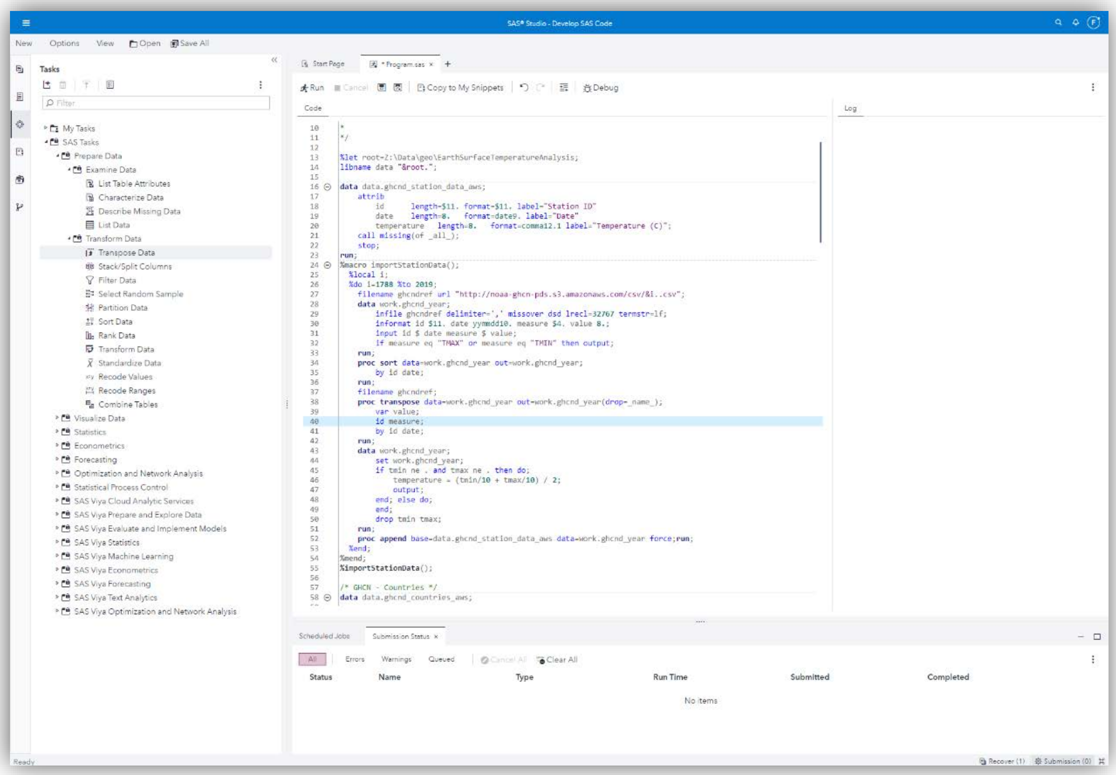
EXAMPLE WORKFLOW

In this paper, the author will analyze Earth's surface temperatures using daily and monthly National Aeronautics and Space Administration (NASA)/GISS data for over 25,000 stations across the globe. We will go through all steps required from initial data collection, preparation, design, and building a full infographic. You will be able to use the same steps in your own design projects.

PREPARATION

Getting access to the data can often be challenging. Whether it's related to the legal permissions required to be able to use the data or simply having data that is not in the right format or structure. For this example, the author uses open data from NASA and National Oceanic and Atmospheric Administration (NOAA) that published global historical temperature data as part of the Global Historical Climatology Network (GHCN). We are also going to include results of the related GISS Surface Temperature Analysis (GISTEMP v4) for hundreds of weather stations around the world.

GHCN data is provided in text files and fixed column layout as described in https://www1.ncdc.noaa.gov/pub/data/ghcn/v4/readme.txt. Since data was provided in such special format, the author decided to make use of the custom code developed in SAS Studio in order to import the data into SAS Viya.



```
*
*/

%let root=Z:\Data\geo\EarthSurfaceTemperatureAnalysis;
libname data "&root.";

data data.ghcnd_station_data_aws;
    attrib
        id          length=$11. format=$11. label="Station ID"
        date        length=8.   format=date9. label="Date"
        temperature   length=8.   format=comma12.1 label="Temperature (C)";
    call missing(of _all_);
    stop;
run;
%macro importStationData();
  %local i;
  %do i=1788 %to 2019;
    filename ghcndref url "http://noaa-ghcn-pds.s3.amazonaws.com/csv/&i..csv";
    data work.ghcnd_year;
        infile ghcndref delimiter=',' missover dsd lrecl=32767 termstr=lf;
        informat id $11. date yymmdd10. measure $4. value 8.;
        input id $ date measure $ value;
        if measure eq "TMAX" or measure eq "TMIN" then output;
    run;
    proc sort data=work.ghcnd_year out=work.ghcnd_year;
        by id date;
    run;
    filename ghcndref;
    proc transpose data=work.ghcnd_year out=work.ghcnd_year(drop=_name_);
        var value;
        id measure;
        by id date;
    run;
    data work.ghcnd_year;
        set work.ghcnd_year;
        if tmin ne . and tmax ne . then do;
            temperature = (tmin/10 + tmax/10) / 2;
            output;
        end; else do;
        end;
        drop tmin tmax;
    run;
    proc append base=data.ghcnd_station_data_aws data=work.ghcnd_year force;run;
  %end;
%mend;
%importStationData();

/* GHCN - Countries */
data data.ghcnd_countries_aws;
```

Figure 1. SAS Studio Code Editor

GISTEMP is provided as compressed NetCDF Files (.NC file extension, regular 2°×2° grid) and the import into SAS requires a conversion to CSV first. The author decided to convert NC files to CSV files using the open-source CDO tool (https://code.mpimet.mpg.de/projects/cdo). The following command was used for the conversion:

```
cdo outputtab,date,lon,lat,value gistemp1200_GHCNv4_ERSSTv5.nc >
gistemp1200_GHCNv4_ERSSTv5.csv
```

The resulting CSV file (which is tab-delimited) can be directly imported into SAS Viya for processing. See Error! Reference source not found. for the full SAS code snippet used here.

EXPLORATION

Before starting with the actual design of our infographics, let’s explore the data and discover interesting insights worthy of sharing in the final report. During the exploration you can validate data quality, discover potential issues, or simply learn about suitable ways of representing information using the best possible visualization.

The data provided is all about historical surface temperature changes across the globe. Some of the questions we are trying to answer include the following:

- Have there been any temperature changes over the years at all?

- How do temperature anomalies compare to the base reference period 1950-1980?
- Are changes the same across the globe? Are some latitude zones or pole regions affected more than others?
- Do monthly mean global surface temperatures increase or decrease?
- Does the Northern or Southern Hemisphere show the same result?
- Do some years show an exceptional difference compared to others?
- How do seasons (e.g., winter/summer) compare across years?
- Are there any hotspots across the globe with higher than average readings?
- Given the 25K+ stations in the inventory, do all provide the same year coverage?

To answer some of these questions and to gain further understanding of the data, let's dive into SAS Visual Analytics and load previously imported data. Note, that global mean temperature anomalies shown in the graphics below are always based on the reference years 1950-1980. I'm starting with the GHCN weather station data (*GHCND_STATIONS* table) to get a first feel where the temperature sensor data is coming from. What better visualization to use as a geographical map in this case? Conveniently station information includes related map coordinates along with station name and elevation.

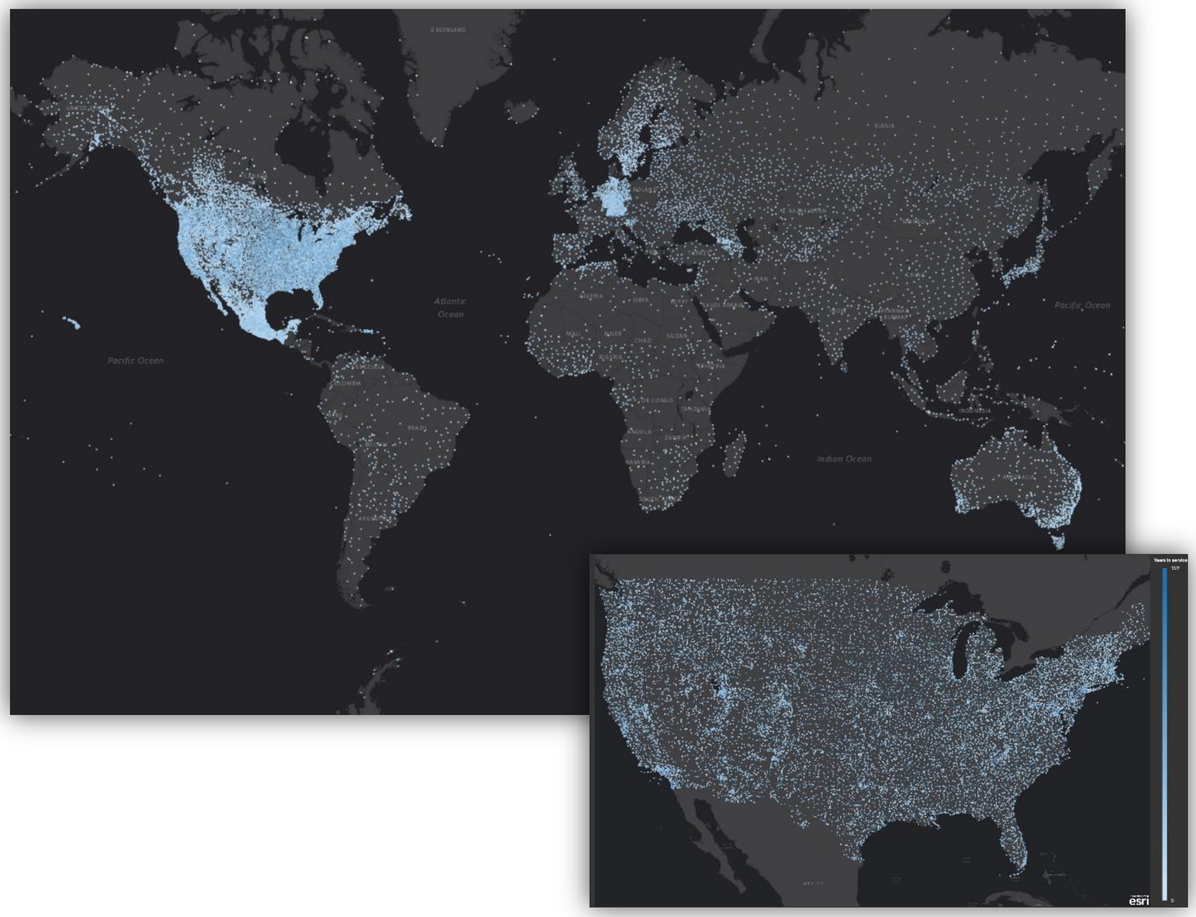


Figure 2. Weather Station Geographical Distribution

The coloring by number of years that a station provided data for nicely shows the distribution of long servicing areas.

Before looking into individual station or region data, let's a have a look at some global measures using the *GISTEMP_MONTHLY_ANOMALY* table that contains monthly global means. Creating a simple time series line plot by year shows:

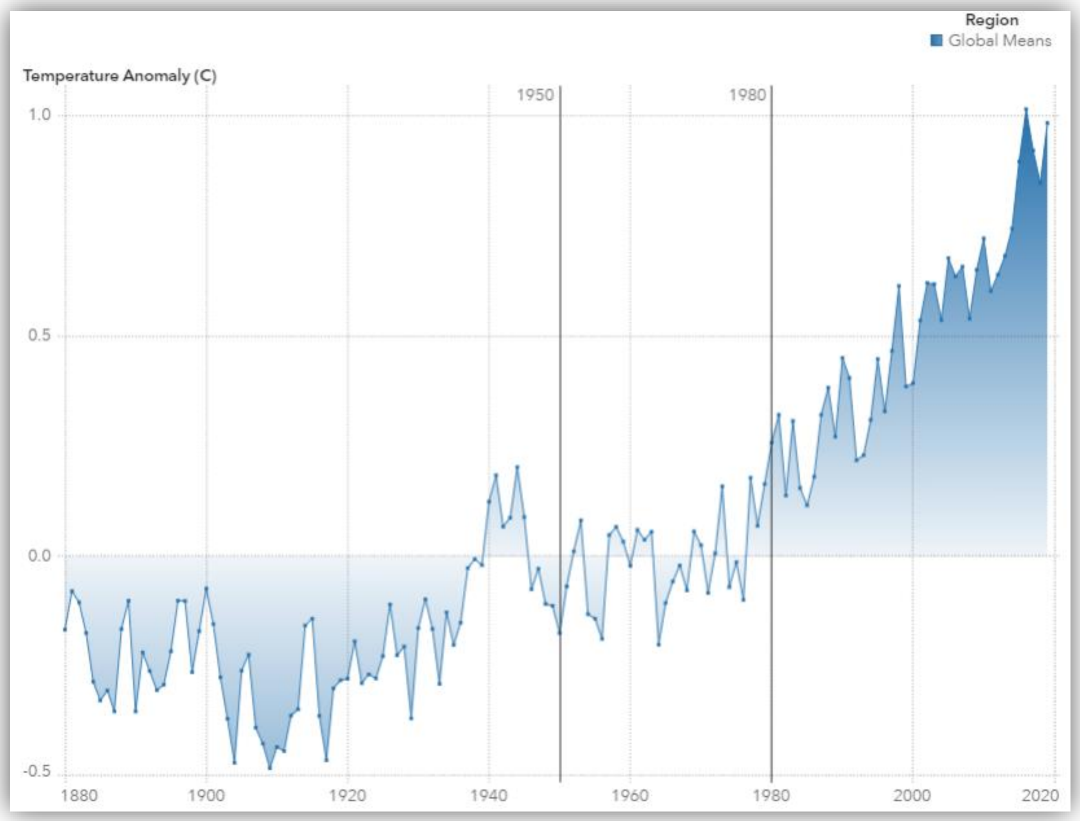


Figure 2. Global –Mean Surface Temperatures Versus 1950 -1980

We can see that overall temperatures increased (a positive anomaly value) compared to the reference years 1950-1980 with especially high readings in recent years. The year 2016 was one of the hottest on record.

Do increase temperatures occur evenly across the globe? Find out by comparing the Northern and Southern Hemispheres:

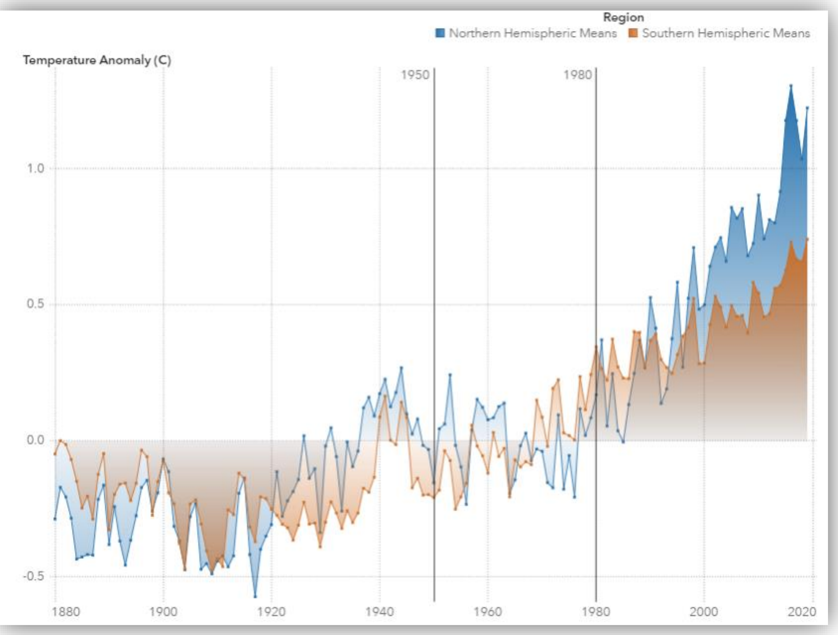


Figure 3. Hemispheric –Means Versus 1950

It seems the Northern Hemisphere shows higher than average readings compared to the base reference years. This could be something to explore further when comparing latitude zone or polar region data.

The data also provides information on a month level and as such we can compare seasonal cycles over the years. The following chart shows the monthly global means by 20-year interval.

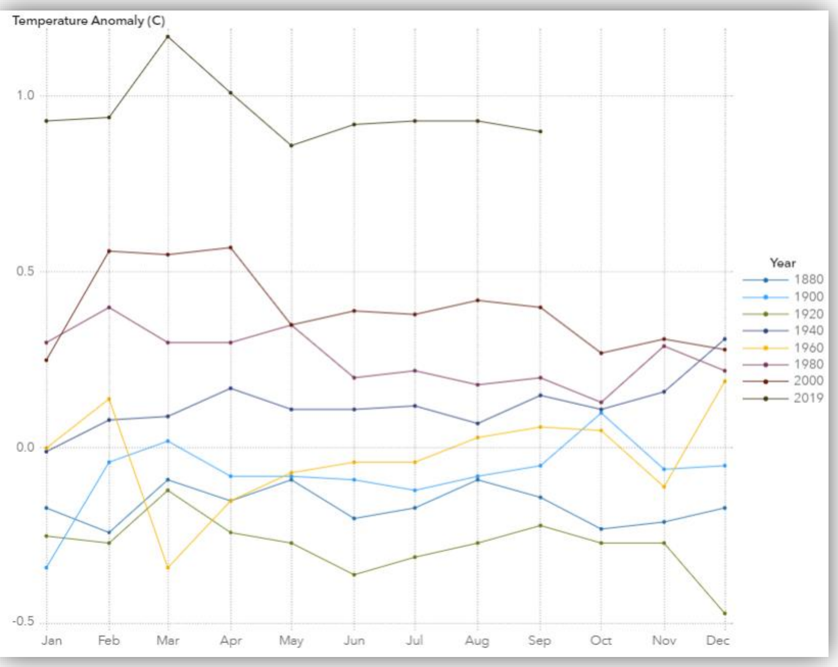


Figure 4. Seasonal Cycles Since 1980

The visualization again confirms higher readings in recent years. It also shows a trend that we typically see higher anomalies beginning the year – likely contributed by the overall higher values for the Northern Hemisphere and related warmer winter season.

Finally, let's look at the gridded surface air temperatures anomaly data, provided by the netCDF files imported earlier. The following contour plot on a geographical map shows the global means for the year 2018.

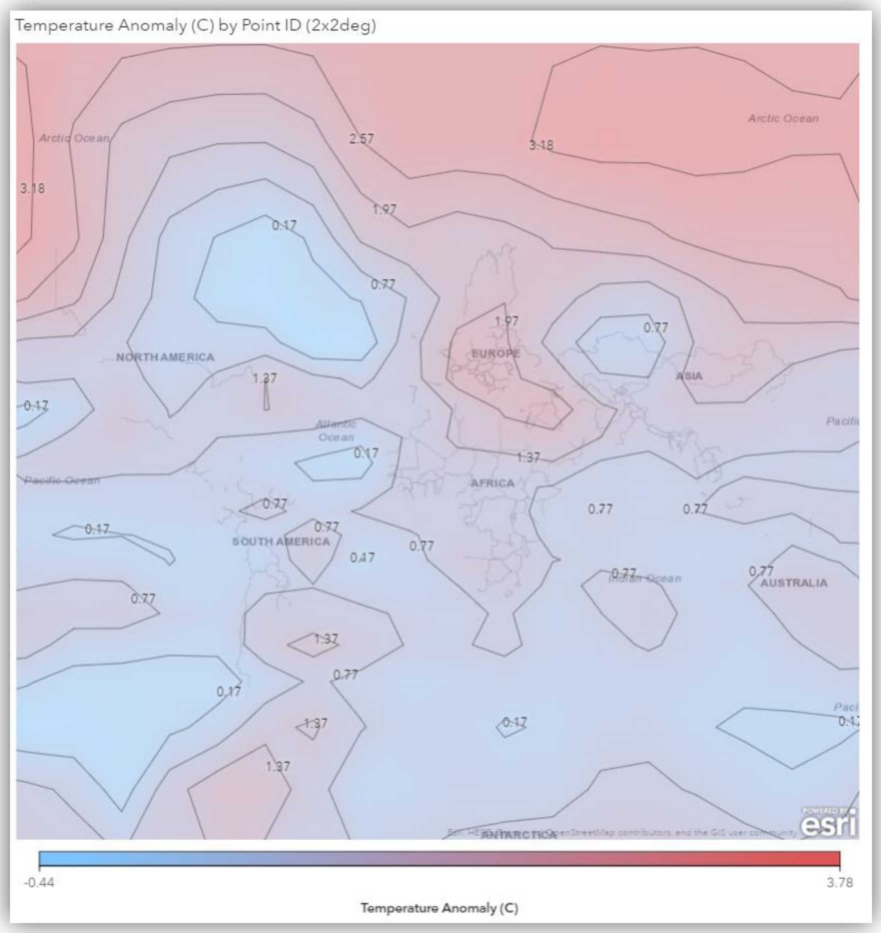


Figure 5. Gridded Surface Air Temperature Anomaly (2018)

We can see hotspots of higher than average temperature changes especially in the Northern Hemisphere near Europe and the polar regions, which is expected given the results of previous visualizations.

Grouping data into latitude zones also confirms this observation:

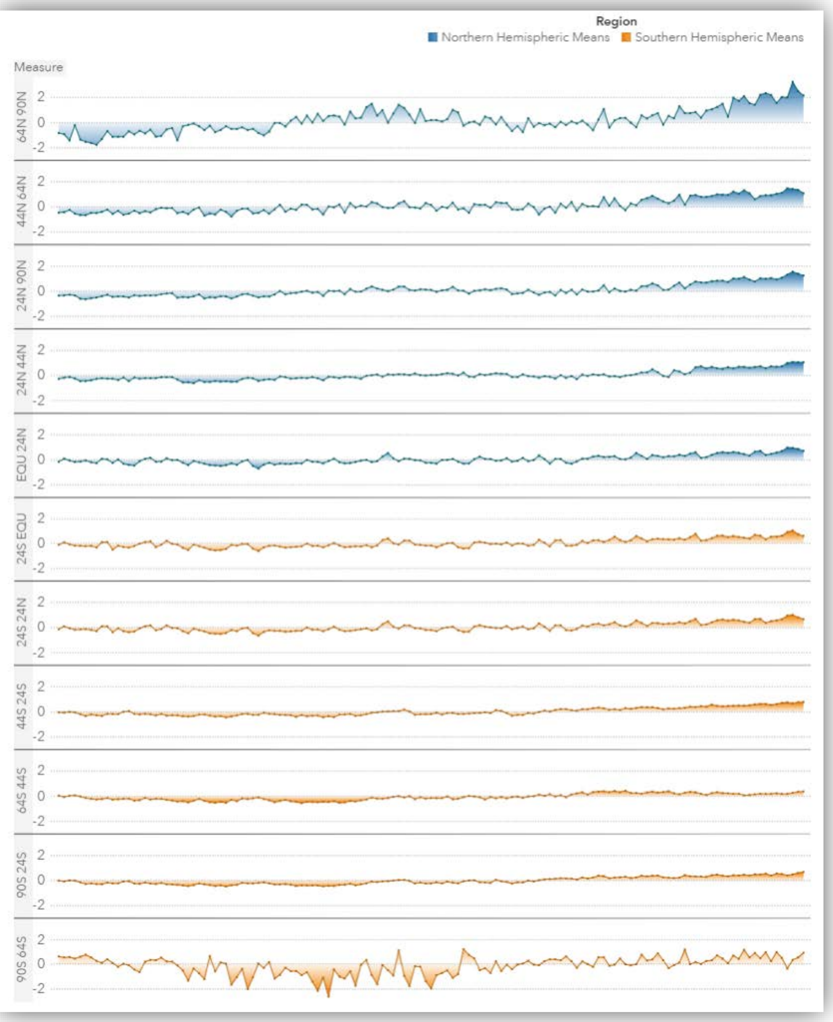


Figure 6. Temperature Anomaly at the Poles

While we could dive into the actual data analysis further and apply things like advanced analytics, this paper focuses on the design and reporting aspect. We collected enough information to understand the data structure, quality, and high-level trends.

DESIGN

Having analyzed the data and gathered insight, we can now continue our journey and tell a story. As report author, the analyst wants to share information learned in the most impactful way. Initial page design mock-ups should reflect our audience who could either be data scientists trying to understand details about our analysis or even the general public given the increased interest and political coverage in recent years.

For the purpose of this paper, the author decided to describe the steps required for a basic infographic covering information about global temperature changes on monthly and yearly levels. While you could start designing your graphics layout in the tool itself, it’s often easier to use a whiteboard or even tools like PowerPoint for the initial mock-up design. This way you don’t get distracted by tool features and avoid being forced into a specific design structure. Figure 7. Initial DesignFigure 7. Initial Design below shows my rough initial design using a PowerPoint slide:

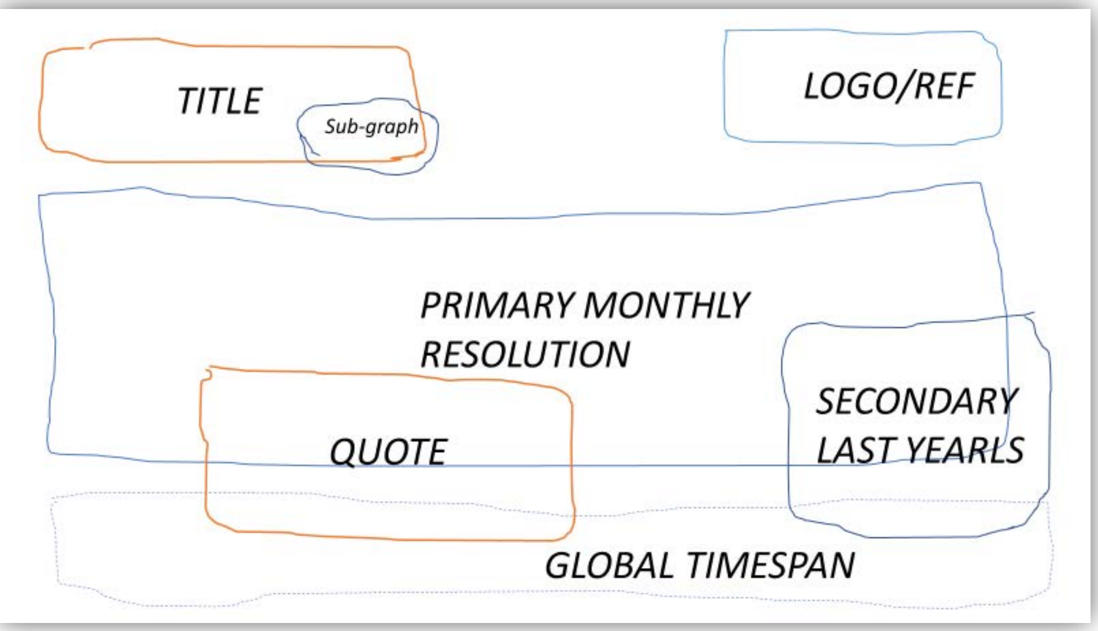


Figure 7. Initial Design

The idea is to have a primary (line) graph that shows the most recent years of monthly global temperature changes in the center and support it by a secondary (bar) chart to the right showing the yearly resolution.

I'm planning to add a third (bar) graph at the bottom showing the entire history of data available, starting with the year 1788. To give the report a more infographic appeal, I'm also going to add a bold title along with a supporting quotation.

While this is a relatively simple infographic, we will be able to cover a few basic design skills during creation. I'm envisioning a darker color palette for the entire infographic, so I will use very dark gray background (color hex code #343433) with mostly white foreground text.

The next step in the design process is to layout the proposed objects in SAS Visual Analytics. As the design requires overlaying objects on top of each other, the first step is to add a special container type supporting it.

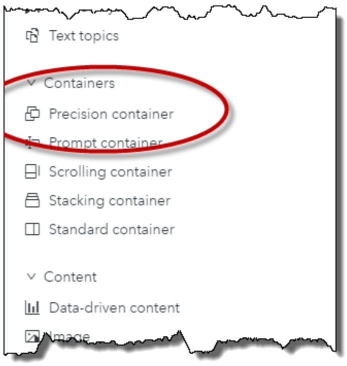


Figure 8. Containers

The Precision container behaves just like a PowerPoint slide and enables you to layout objects freely on a grid. You can control object sizes within such container either by drag and drop or by adjusting size and position in the Objects pane.

I'm also going to adjust the report size to a fixed pixel ratio (1300x980px). This enables me to design the graphics regardless of final screen resolution and ensures that objects are always sized and aligned according to my settings. While you could also use the default responsive layout, some object components (e.g., text) do not scale as well as graph objects and might cause unexpected scrollbars or rendering. You can set a fixed report size in the Options pane using the report-level options.

BUILD

Let's add the proposed graph objects into our container and build the actual report page. You could either make use of the ready-to-use page templates or start with a blank canvas in SAS Visual Analytics. None of the templates seem to fit in this case, so I'm going to drag and drop objects into the precision container and align them per my initial design. Figure 9. Initial shows the graphs rendered as I expected with the sample data.

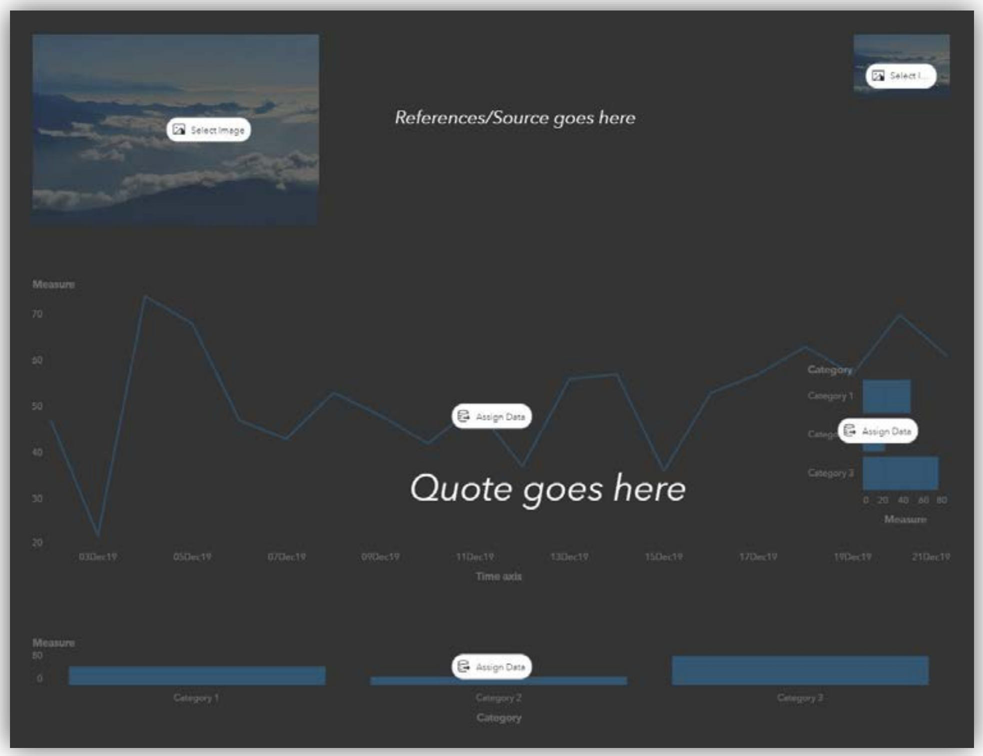


Figure 9. Initial Design in SAS Visual Analytics

While building your report, it's important to keep track of the objects that you have added. An excellent tool is the Outline pane. It shows the hierarchical structure of a report, providing access to object-specific actions. The Outline pane enables you to interact with objects even though you might not be able to see them on the canvas (for example, one object might be overlaid by another object). Using the right-click menu will provide access to object options. For better reference, you might consider renaming objects to better reflect what they represent.

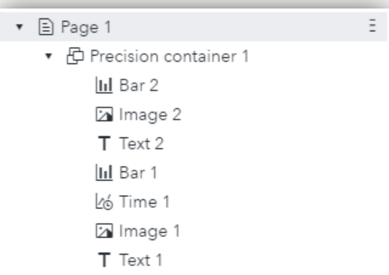

Page 1
Precision container 1
Bar 2
Image 2
Text 2
Bar 1
Time 1
Image 1
Text 1

Figure 10. Default Object Names

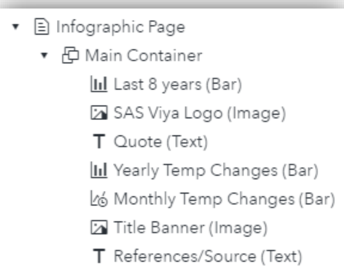

Infographic Page
Main Container
Last 8 years (Bar)
SAS Viya Logo (Image)
Quote (Text)
Yearly Temp Changes (Bar)
Monthly Temp Changes (Bar)
Title Banner (Image)
References/Source (Text)

Figure 11. Customized Object Names

Now let's bring graphics to life by adding the *GHCNM_STATION_DATA_AWS* data source. This table contains the aggregated monthly temperature (in degrees Celsius) and temperature changes (compared to the base period) since data recording started. Let's start with the bottom bar chart, which should visualize the temperature changes since 1788. Given the chart will show data on year level, I'm going to create a copy of the *Date* variable by right-clicking and selecting Duplicate, and then selecting the *Year* format. Assigning this newly created variable *Year* to the bar chart as shown in Figure 12 – Unformatted provides a first sobering look of the visual.

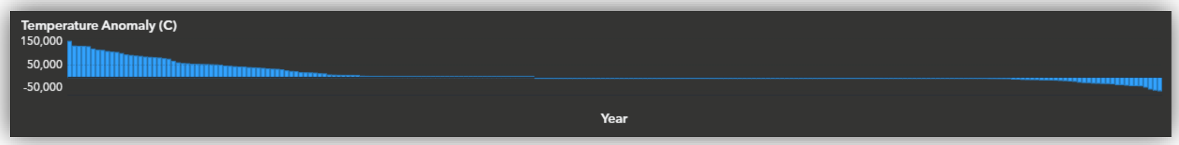
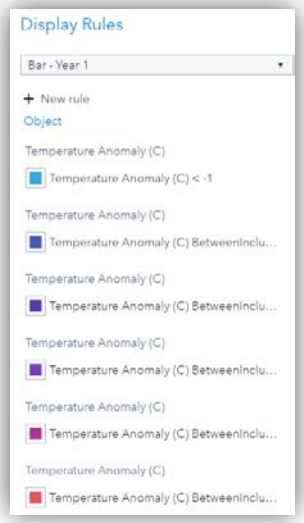


Figure 12 – Unformatted Data from the GHCNM_STATION_DATA_AWS Table

To get the rendering I want, I’m going to apply the following option changes:

- Sort date ascending by year
- Change the aggregation for variable *Temperature Anomaly (C)* to average
- Remove the graph grid lines
- Remove the X- and Y-axis labels, lines, and tick values
- Add two text objects to the left and the right of the graph showing the start and end years
- Add display rules to color code the bars given their temperature anomaly. I used the following color codes and rules, as shown in Figure 13. Display Rules:
    - Value > 1: #C8346F
    - 0.5 < Value < 1: #B23196
    - 0 < Value < 0.5: #7B3DBA
    - -0.5 < Value < 0: #5839B5
    - -1 value < -0.5: #425CBC
    - Value < -1: #2AAFE0



Display Rules

Bar - Year 1 ▾

+ New rule

Object

Temperature Anomaly (C)

🟦 Temperature Anomaly (C) < -1

Temperature Anomaly (C)

🟦 Temperature Anomaly (C) BetweenInclu...

Temperature Anomaly (C)

🟪 Temperature Anomaly (C) BetweenInclu...

Temperature Anomaly (C)

🟪 Temperature Anomaly (C) BetweenInclu...

Temperature Anomaly (C)

🟪 Temperature Anomaly (C) BetweenInclu...

Temperature Anomaly (C)

🟥 Temperature Anomaly (C) BetweenInclu...

Figure 13. Display Rules

Figure 14. shows the resulting visual. The additional coloring of bars gives a much better visual support for higher and lower temperature changes over the years.

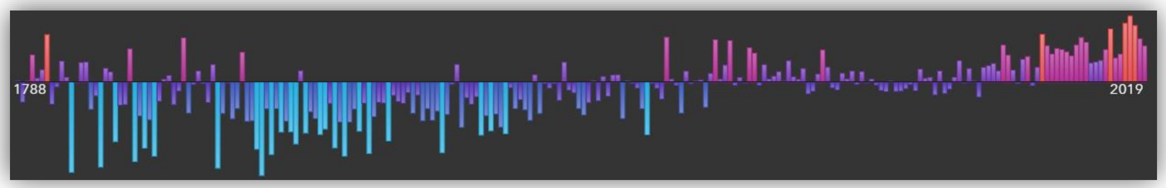


Figure 14. Resulting Visual with Color and Additional Options

Let’s investigate the primary center graph now. This visualization shows the seasonal variations of temperature changes. Using a line chart here would be great due to the nature of time series data. Now, like before, I’m creating a copy of the *Date* variable and specifying the *Month* format. Adding this newly created variable and group by *Year* to the center chart shows a rather cluttered first glance of the data:

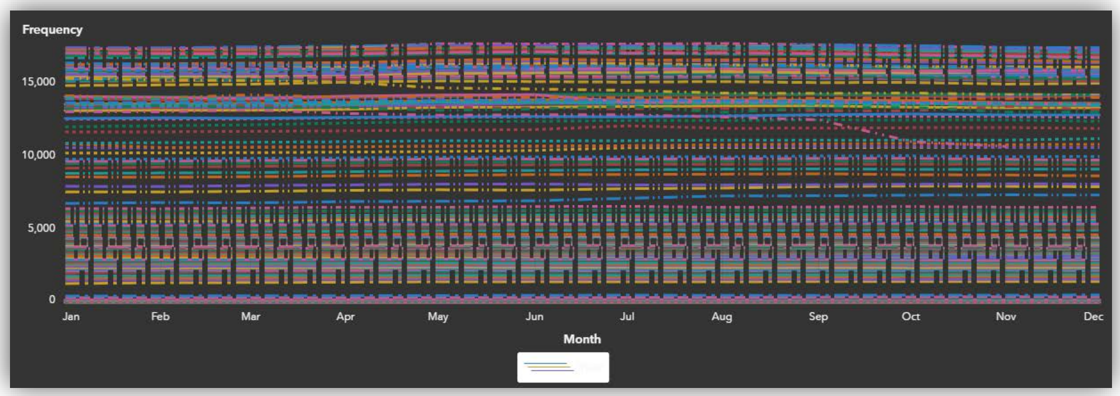


Figure 16. The Primary Center Graph Before Revisions

Given the high number of years, I'm going to reduce data for rendering and showcase the last 20 years. However, even 20 years is a relatively high number of lines to be rendering on top of each other. I'm using additional display rules to highlight only the last 8 years. I'm going to apply the following changes to the graph:

- Add a filter for Year >= 2012
- Remove the graph grid lines
- Turn on markers (size = 3)
- Change the line thickness to 2
- Show the data labels (text style 12, bold)
- X-axis options
    - Remove the axis label
- Y-axis options
    - Remove the axis label
    - Set a fixed minimum = -1
    - Set a fixed maximum = 3
    - Add a reference line at 0.94 (label = "20th Century Average")
- Line/Marker colors:
    - 2012: #97368E
    - 2013: #5839B5
    - 2014: #6538AC
    - 2015: #E03361
    - 2016: #DD5757
    - 2017: #C13473
    - 2018: #97368E
    - 2019: #81379B
- Turn the legend visibility off

Applying above changes renders the graph as follow:

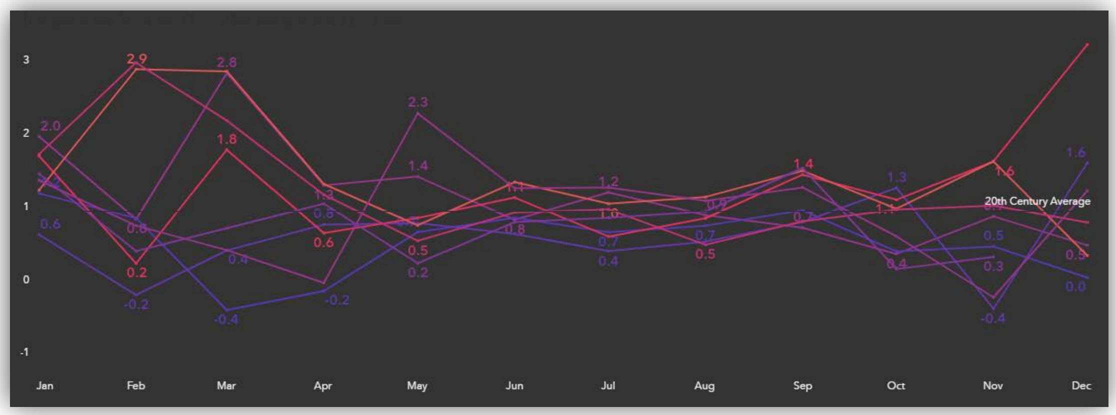


Figure 17. Monthly Temperature Changes from 2012 - 2019

Additionally, I want to highlight the fact that we have many more years of data, so I'm going to add back the previous filtered-out 12 years. Now, for better visualization and not taking the focus away from the main graph, I want those years to fade into the background and barely be visible. To achieve that effect, I'm creating a copy of the above graph (right-click, and select Duplicate) and making the following changes:

- Change all the Line/Marker colors to a dark gray color (hex color 3D4550)
- Reduce line thickness to 1
- Remove all data labels
- Change the layout position and the size to be exactly on top of the original graph
- Move the new graph to the back so that the original graph overlays on top
- Remove the X- and Y-axis labels and the tick values

The result is a nicely faded graph behind the primary visualization:

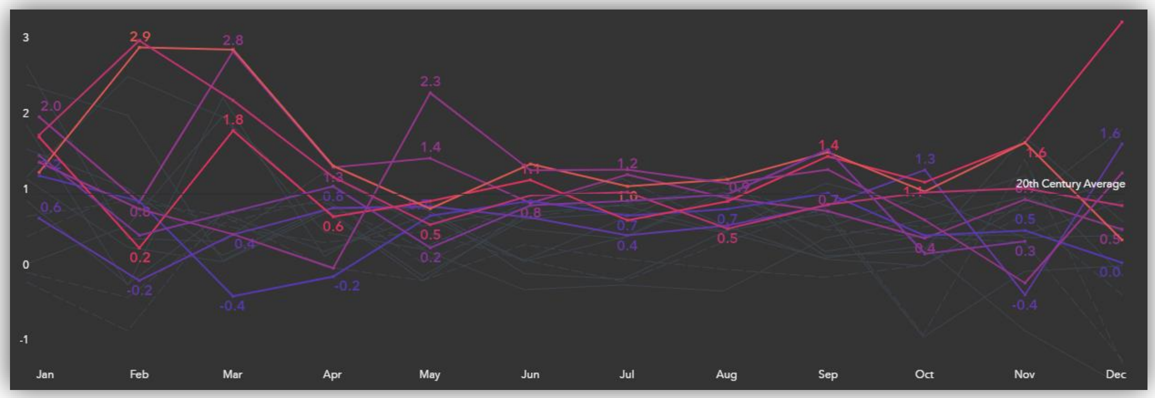


Figure 18. Monthly Temperature Changes with a Background Graph

Lastly, let's work on the overlaying the bar chart to the right of my initial design. As you have noticed, we removed the legend in one of the previous steps when tweaking the line chart properties. Now, every graph needs a legend if the colors or the symbols aren't self-explaining. I removed the original legend as I wanted to replace it with a richer bar chart that shows the yearly aggregated values. Special care is required here to synchronize the colors between the line and the bar chart.

The bar chart uses the previously created *Year* variable as category and the temperature change as the bar height. As before, the initial rendering isn't too appealing at first:

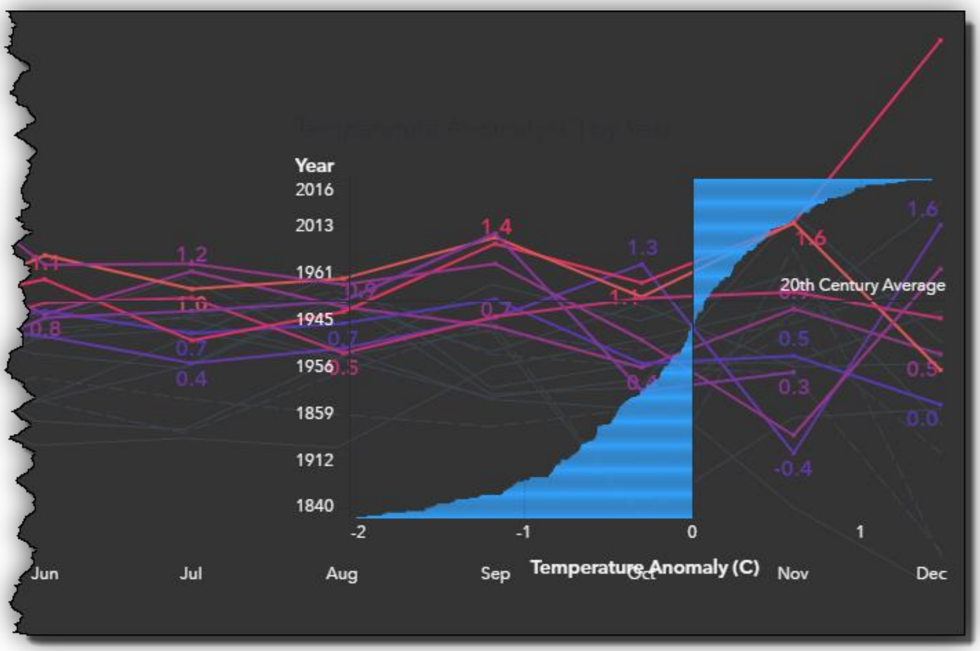


Figure 19. Initial Bar Chart Rendering

To improve the rendering, I'm going to apply the following options:

- Add a filter for Year >= 2012
- Remove the object title
- Reduce the object size to 15% width and 25% height and align to the right
- Remove graph grid lines
- Add segment labels (text size 10)
- Y-Axis options
  - Remove labels
  - Remove line
- X-Axis options
  - Remove labels
  - Remove line
  - Remove tick values
- Line/Marker colors:
  - 2012: 97368E
  - 2013: 5839B5
  - 2014: 6538AC
  - 2015: E03361
  - 2016: DD5757
  - 2017: C13473
  - 2018: 97368E
  - 2019: 81379B
- Increase the transparency to 20%

The above changes help readers to identify years with high versus low temperature changes in the following graph:

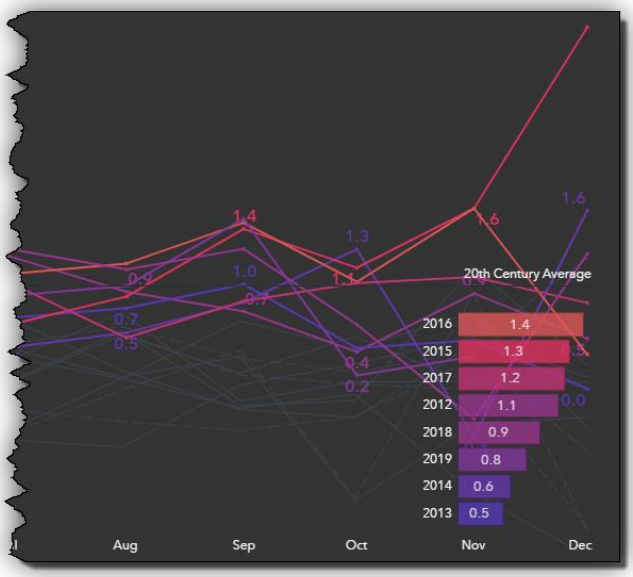


Figure 20. The Last Eight Years of Temperature Changes

FINAL TWEAKS

The final tweaks include adding the top left banner, as well as adding the SAS Viya logo along with some additional text about the data sources that are used. I used PowerPoint to create the faded background banner image that includes the 'tipping point' graphic. If you don't need a fancy banner, you might want to use the built-in text objects that allow rich text formatting.

To further support the rising temperature changes and tipping of the related graphics, I overlaid the banner image with an additional line chart that shows temperature changes by year. While this is optional, it gives it a nice supporting touch.

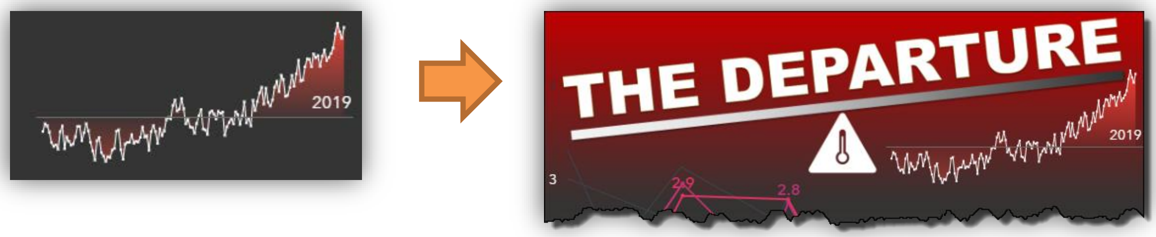
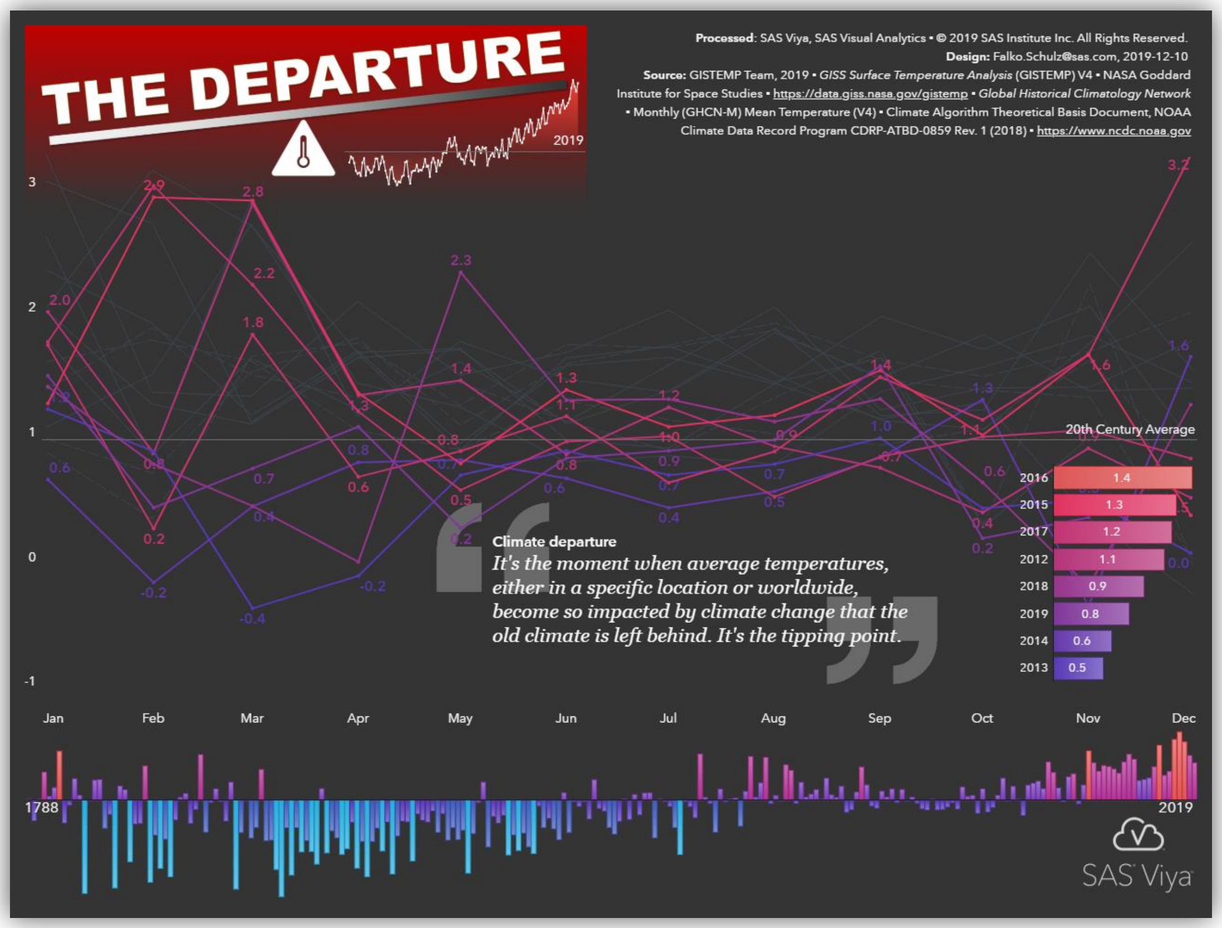


Figure 21. The Banner for the Report

The final version of the infographic also includes a quotation to further support the message about rising temperatures globally. Note that I used a custom image for the double quotation marks which nicely visualizes the text as a quotation. You may also use the text object here with special formatting.

The final infographic renders as follows:



Processed: SAS Viya, SAS Visual Analytics • © 2019 SAS Institute Inc. All Rights Reserved.
Design: Falko.Schulz@sas.com, 2019-12-10
Source: GISTEMP Team, 2019 • *GISS Surface Temperature Analysis* (GISTEMP) V4 • NASA Goddard Institute for Space Studies • https://data.giss.nasa.gov/gistemp • *Global Historical Climatology Network* • Monthly (GHCN-M) Mean Temperature (V4) • Climate Algorithm Theoretical Basis Document, NOAA Climate Data Record Program CDRP-ATBD-0859 Rev. 1 (2018) • https://www.ncdc.noaa.gov

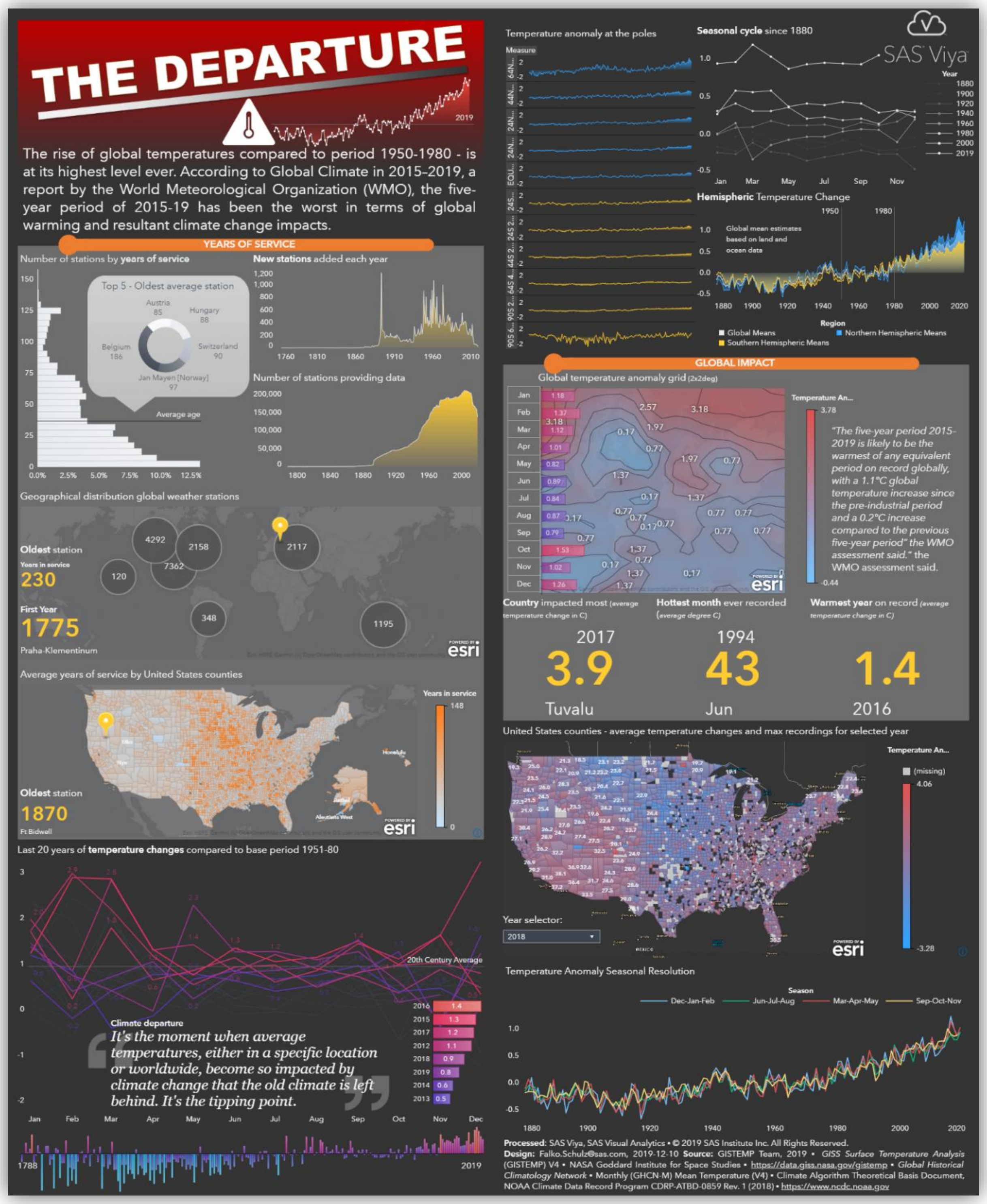
Climate departure
It's the moment when average temperatures, either in a specific location or worldwide, become so impacted by climate change that the old climate is left behind. It's the tipping point.

Figure 22. The Final Infographic

You might want to view the speed-art compilation of the above infographic segment and see how the design process is done in the application: https://youtu.be/RiFoAH-XWiY.

Taking this a step further, the author also decided to include the above segment into an even larger graphic covering all aspects of surface air temperature data and related analysis (bottom left). The infographic shows the distribution of weather stations globally and their years of service. The viewer will also learn about temperature anomalies at the poles and general impact of changes around the world.

Breaking down complex infographics into smaller dashboards is a common practice and makes the development and design process easier. It also helps to structure such large-scale posters into segments and topic of interest. In this example, different background gradients help consumers to differentiate between analysis results of weather stations, as well as seasonal and global air temperature measurements.



THE DEPARTURE

The rise of global temperatures compared to period 1950-1980 - is at its highest level ever. According to Global Climate in 2015-2019, a report by the World Meteorological Organization (WMO), the five-year period of 2015-19 has been the worst in terms of global warming and resultant climate change impacts.

YEARS OF SERVICE

GLOBAL IMPACT

"The five-year period 2015-2019 is likely to be the warmest of any equivalent period on record globally, with a 1.1°C global temperature increase since the pre-industrial period and a 0.2°C increase compared to the previous five-year period" the WMO assessment said." the WMO assessment said.

Climate departure

It's the moment when average temperatures, either in a specific location or worldwide, become so impacted by climate change that the old climate is left behind. It's the tipping point.

Processed: SAS Viya, SAS Visual Analytics • © 2019 SAS Institute Inc. All Rights Reserved.
Design: Falko.Schulz@sas.com, 2019-12-10 **Source:** GISTEMP Team, 2019 • *GISS Surface Temperature Analysis* (GISTEMP) V4 • NASA Goddard Institute for Space Studies • https://data.giss.nasa.gov/gistemp • *Global Historical Climatology Network* • Monthly (GHCN-M) Mean Temperature (V4) • Climate Algorithm Theoretical Basis Document, NOAA Climate Data Record Program CDRP-ATBD-0859 Rev. 1 (2018) • https://www.ncdc.noaa.gov

2019
Number of stations by years of service
Top 5 - Oldest average station
Austria 85
Hungary 88
Belgium 186
Switzerland 90
Jan Mayen (Norway) 97
Average age
New stations added each year
Number of stations providing data
Geographical distribution global weather stations
Oldest station
Years in service
230
First Year
1775
Praha-Klementinum
4292
2158
7362
120
2117
348
1195
esri
Average years of service by United States counties
Years in service
148
0
Oldest station
1870
Ft Bidwell
Last 20 years of temperature changes compared to base period 1951-80
20th Century Average
2016 1.4
2015 1.3
2017 1.2
2012 1.1
2018 0.9
2019 0.8
2014 0.6
2013 0.5
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1788
2019
Temperature anomaly at the poles
Measure
Seasonal cycle since 1880
SAS Viya
Year
1880 1900 1920 1940 1960 1980 2000 2019
Jan Mar May Jul Sep Nov
Hemispheric Temperature Change
1950
1980
Global mean estimates based on land and ocean data
Region
Global Means
Northern Hemispheric Means
Southern Hemispheric Means
Global temperature anomaly grid (2x2deg)
Temperature An...
3.78
-0.44
Country impacted most (average temperature change in C)
Hottest month ever recorded (average degree C)
Warmest year on record (average temperature change in C)
2017
3.9
Tuvalu
1994
43
Jun
1.4
2016
United States counties - average temperature changes and max recordings for selected year
Temperature An...
(missing)
4.06
-3.28
Year selector:
2018
Temperature Anomaly Seasonal Resolution
Season
Dec-Jan-Feb
Jun-Jul-Aug
Mar-Apr-May
Sep-Oct-Nov
1880 1900 1920 1940 1960 1980 2000 2020

Figure 23. The Full-Size Paper Infographic

CONCLUSION

SAS Visual Analytics is a powerful and feature rich toolset for any data scientist or business analyst. In this paper, the author outlined the best practices and the general steps needed in order to achieve impactful visualizations. Along with intuitive user design and drag-and drop-controls in a self-service interface, creating dashboards and infographics will quickly become the norm for engaging reporting.

So, use your imagination and make your analytical reports even more appealing and interesting. Don't forget to share your examples in the SAS Visual Analytics gallery!

REFERENCES

Rouse, Margaret. "What Is a Dashboard?" WhatIs.com, Last updated September 2005. Accessed February 20, 2018. http://searchcio.techtarget.com/definition/dashboard

"Infographic." Wikipedia. Accessed February 20, 2018. https://en.wikipedia.org/wiki/Infographic

GISTEMP Team, 2019: GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies. Dataset accessed November 21, 2019. https://data.giss.nasa.gov/gistemp

Lenssen, N., G. Schmidt, J. Hansen, M. Menne,A. Persin,R. Ruedy, and D. Zyss, 2019: Improvements in the GISTEMP uncertainty model. J. Geophys. Res. Atmos., 124, no. 12, 6307-6326, doi:10.1029/2018JD029522

Global Historical Climatology Network - Monthly (GHCN-M) Mean Temperature (Version 4) - Climate Algorithm Theoretical Basis Document, NOAA Climate Data Record Program CDRP-ATBD-0859 Rev. 1 (2018). Dataset accessed November 21, 2019. https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-monthly-version-4

RECOMMENDED READING

- Murphy, Travis. 2018. *Infographics Powered by SAS: Data Visualization Techniques for Business Reporting*. Cary, NC: SAS Institute Inc.
- Schulz, Falko. 2017. "How to Design a Meteorite Infographic Using NASA Data and SAS." SAS Voices (blog). SAS Institute Inc. http://blogs.sas.com/content/sascom/2017/03/28/design-meteorite-infographic-using-nasa-data-sas. Last modified March 28, 2017.
- SAS Institute Inc. 2018. SAS Visual Analytics: Designing Reports. https://www.sas.com/beautifulreports.
- SAS Institute Inc. 2019. SAS Visual Analytics: Gallery. https://communities.sas.com/t5/SAS-Visual-Analytics-Gallery/tkb-p/vagallery.

APPENDIX A

The code snippet below shows the detail steps taken to import GHCN and GISTEMP data into SAS data sets. Such code can be scheduled to run in batch and provide a consistent import and update of data files.

```
/**
* Copyright (c) 2019, SAS Institute Inc., Cary, NC, USA, All Rights Reserved
*
* import.sas; author: falko.schulz@sas.com, 2019
*
* This code snippet imports:
*
*   - GHCN: https://www1.ncdc.noaa.gov/pub/data/ghcn/v4
*   - GISTEMP: https://data.giss.nasa.gov/gistemp
*
*/

%let root=C:\SAS\EarthSurfaceTemperatureAnalysis;
libname data "&root.";

/* GHCN - Flags */
data data.ghcnm_flags;
      infile "&root.\GHCN\ghcnm-flags.txt" missover dsd lrecl=32767;
      informat flag $1. source $11. sourcedesc $200.;
      format flag $1. source $11. sourcedesc $200.;
      input flag 1-1 source 6-17 @18 sourcedesc;
run;

/* GHCN - Countries */
data data.ghcnm_countries;
      infile "&root.\GHCN\ghcnm-countries.txt" missover dsd lrecl=32767;
      informat country_abbr $2. country $50.;
      format country_abbr $2. country $50.;
      input country_abbr @2 +1 country $;
run;

/* GHCN - Weather stations */
data data.ghcnm_stations;
      infile "&root.\GHCN\ghcnm.tavg.v4.0.1.20191107.qfe.inv" missover dsd
lrecl=32767;
      informat id $11. latitude 8. longitude 8. elevation 8. name $50.;
      format id $11. latitude best12. longitude  best12. elevation best12.
name $50.;
      label id="Station Id" latitude="Latitude" longitude="Longitude"
elevation="Elevation (m)" name="Station Name";
      input id @13 latitude 13-20 longitude 22-30 elevation 32-37 name;

      name=compress(tranwrd(name, "**", " "));
      name=tranwrd(name, "_", " ");
      name=propcase(name);

      if (elevation eq -999) then elevation = .;
run;

data work.ghcnm_station_data;
      infile "&root.\GHCN\ghcnm.tavg.v4.0.1.20191107.qfe.dat" missover dsd
lrecl=32767;
```

```
      informat id $11. year 8. element $4.
                  value1 5. dmflag1 $1. qcflag1 $1. dsflag1 $1.
                  value2 8. dmflag2 $1. qcflag2 $1. dsflag2 $1.
                  value3 8. dmflag3 $1. qcflag3 $1. dsflag3 $1.
                  value4 8. dmflag4 $1. qcflag4 $1. dsflag4 $1.
                  value5 8. dmflag5 $1. qcflag5 $1. dsflag5 $1.
                  value6 8. dmflag6 $1. qcflag6 $1. dsflag6 $1.
                  value7 8. dmflag7 $1. qcflag7 $1. dsflag7 $1.
                  value8 8. dmflag8 $1. qcflag8 $1. dsflag8 $1.
                  value9 8. dmflag9 $1. qcflag9 $1. dsflag9 $1.
                  value10 8. dmflag10 $1. qcflag10 $1. dsflag10 $1.
                  value11 8. dmflag11 $1. qcflag11 $1. dsflag11 $1.
                  value12 8. dmflag12 $1. qcflag12 $1. dsflag12 $1.;
      format      id $11. year 8. element $4.
                  value1 comma12.2 dmflag1 $1. qcflag1 $1. dsflag1 $1.
                  value2 comma12.2 dmflag2 $1. qcflag2 $1. dsflag2 $1.
                  value3 comma12.2 dmflag3 $1. qcflag3 $1. dsflag3 $1.
                  value4 comma12.2 dmflag4 $1. qcflag4 $1. dsflag4 $1.
                  value5 comma12.2 dmflag5 $1. qcflag5 $1. dsflag5 $1.
                  value6 comma12.2 dmflag6 $1. qcflag6 $1. dsflag6 $1.
                  value7 comma12.2 dmflag7 $1. qcflag7 $1. dsflag7 $1.
                  value8 comma12.2 dmflag8 $1. qcflag8 $1. dsflag8 $1.
                  value9 comma12.2 dmflag9 $1. qcflag9 $1. dsflag9 $1.
                  value10 comma12.2 dmflag10 $1. qcflag10 $1. dsflag10 $1.
                  value11 comma12.2 dmflag11 $1. qcflag11 $1. dsflag11 $1.
                  value12 comma12.2 dmflag12 $1. qcflag12 $1. dsflag12 $1.;

      input       id 1-11 year 12-15 element 16-19
                  value1 20-24 dmflag1 25-25 qcflag1 26-26 dsflag1 27-27
                  value2 28-32 dmflag2 33-33 qcflag2 34-34 dsflag2 35-35
                  value3 36-40 dmflag3 41-41 qcflag3 42-42 dsflag3 43-43
                  value4 44-48 dmflag4 49-49 qcflag4 50-50 dsflag4 51-51
                  value5 52-56 dmflag5 57-57 qcflag5 58-58 dsflag5 59-59
                  value6 60-64 dmflag6 65-65 qcflag6 66-66 dsflag6 67-67
                  value7 68-72 dmflag7 73-73 qcflag7 74-74 dsflag7 75-75
                  value8 76-80 dmflag8 81-81 qcflag8 82-82 dsflag8 83-83
                  value9 84-88 dmflag9 89-89 qcflag9 90-90 dsflag9 91-91
                  value10 92-96 dmflag10 97-97 qcflag10 98-98 dsflag10 99-99
                  value11 100-104 dmflag11 105-105 qcflag11 106-106 dsflag11
107-107
                  value12 108-112 dmflag12 113-113 qcflag12 114-114 dsflag12
115-115;

      if (value1 eq -9999) then value1 = .; else value1 = value1 / 100;
      if (value2 eq -9999) then value2 = .; else value2 = value2 / 100;
      if (value3 eq -9999) then value3 = .; else value3 = value3 / 100;
      if (value4 eq -9999) then value4 = .; else value4 = value4 / 100;
      if (value5 eq -9999) then value5 = .; else value5 = value5 / 100;
      if (value6 eq -9999) then value6 = .; else value6 = value6 / 100;
      if (value7 eq -9999) then value7 = .; else value7 = value7 / 100;
      if (value8 eq -9999) then value8 = .; else value8 = value8 / 100;
      if (value9 eq -9999) then value9 = .; else value9 = value9 / 100;
      if (value10 eq -9999) then value10 = .; else value10 = value10 / 100;
      if (value11 eq -9999) then value11 = .; else value11 = value11 / 100;
      if (value12 eq -9999) then value12 = .; else value12 = value12 / 100;
run;
```

```
proc transpose data=work.ghcnm_station_data
out=work.ghcnm_station_data_temp(rename=(col1=temperature));
      var value1-value12;
      by id year element;
run;
proc transpose data=work.ghcnm_station_data
out=work.ghcnm_station_data_dm(rename=(col1=dmflag));
      var dmflag1-dmflag12;
      by id year element;
run;
proc transpose data=work.ghcnm_station_data
out=work.ghcnm_station_data_qc(rename=(col1=qcflag));
      var qcflag1-qcflag12;
      by id year element;
run;
proc transpose data=work.ghcnm_station_data
out=work.ghcnm_station_data_ds(rename=(col1=dsflag));
      var dsflag1-dsflag12;
      by id year element;
run;

data data.ghcnm_station_data;
   merge work.ghcnm_station_data_temp work.ghcnm_station_data_dm
work.ghcnm_station_data_qc work.ghcnm_station_data_ds;
   by id year element;
run;

proc sql noprint;
      create table data.ghcnm_station_data(compress=yes) as
      select      a.id,
                  c.country_abbr as country label="Country",
                  mdy(input(compress(a._name_,"dsflag"), 8.),1,a.year) as
date format=monyy5. label="Date",
                  a.temperature label="Temperature (C)",
                  a.dmflag label="Data Measurement Flag",
                  a.qcflag label="Quality Control Flag",
                  a.dsflag label="Data Source Flag",
                  b.latitude,
                  b.longitude,
                  b.elevation,
                  propcase(tranwrd(b.name, "_", " ")) as name label="Station
Name" length=50
      from  data.ghcnm_station_data a
                  left join
                  data.ghcnm_stations b
                  on a.id eq b.id
                  left join
                  data.ghcnm_countries c
                  on substring(a.id from 1 for 2) eq c.country_abbr
;run;quit;

/* GHCN - Calculate the temperature averages between 1951 and 1980 for each
station */
proc sql;
      create table work.temp_1951_1980_mean as
      select      id,
                  month(date) as month,
                  avg(temperature) as temperature
      from  data.ghcnm_station_data
      where year(date) <= 1980
      group by id, calculated month
      order   by id, calculated month
;quit;run;

/* GHCN - Final weather stations data set containing temperature and anomaly
*/
proc sql noprint;
      create table data.ghcnm_station_data(compress=yes) as
      select      a.*,
                  a.temperature - b.temperature as temp_anomaly
format=comma12.2 label="Temperature Anomaly (C)"
      from  data.ghcnm_station_data a
                  left join
                  work.temp_1951_1980_mean b
                  on a.id eq b.id and month(a.date) eq b.month
;run;quit;


/* GISTEMP - Import tabbed data using fixed column layout */
data data.gistemp_temp_anomaly_gridded;
      infile "&root.\gistemp\NC2CSV\gistemp1200_GHCNv4_ERSSTv5.csv" missover
dsd lrecl=32767;
      length pointid 8. date 8.;
      informat tmp $10. longitude 8. latitude 8. temp_anomaly 8.;
      format date monyy5. longitude 8. latitude 8. temp_anomaly comma12.2;
      label date="Date" pointid="Point ID (2x2deg)" longitude="Longitude
(Point)" latitude="Latitude (Point)" temp_anomaly="Temperature Anomaly (C)";
      input tmp 1-11 longitude 12-19 latitude 20-26 temp_anomaly 27-35;
      pointid = _n_ - 1;
      date = input(tmp,yymmdd10.);

      if (temp_anomaly eq 32767) then temp_anomaly = .;
      if (temp_anomaly ne . and date ge "01JAN1980"d) then output;
      drop tmp;
run;

/*  GISTEMP - Global Means */
proc import datafile="&root.\gistemp\GLB.Ts+dSST.csv"
out=work.gistemp_global_means dbms=csv replace;getnames=yes;run;
proc transpose data=work.gistemp_global_means
out=work.gistemp_global_month(rename=(col1=temp_anomaly _name_=month));
      var jan feb mar apr may jun jul aug sep oct nov dec; by year;
run;
data data.gistemp_monthly_anomaly;
      length date 8.;format date monyy5.; label date="Date";
      length type $50.; type="Global Means"; label type="Region";
      set gistemp_global_month;
      date = input(compress(month || put(year,4.)), monyy7.);
      drop year month;
      label temp_anomaly="Temperature Anomaly (C)";
run;
proc transpose data=work.gistemp_global_means
out=work.gistemp_global_season(rename=(col1=temp_anomaly _name_=measure));
      var j_d d_n djf mam jja son; by year;
run;
data data.gistemp_season_anomaly;
```

```
      set work.gistemp_global_season;
      length type $50.; type="Global Means"; label type="Region";
      label measure="Measure";
      label temp_anomaly="Temperature Anomaly (C)";
run;

/*  GISTEMP - Northern Hemispheric Means */
proc import datafile="&root.\gistemp\NH.Ts+dSST.csv"
out=work.gistemp_nh_means dbms=csv replace;getnames=yes;run;
proc transpose data=work.gistemp_nh_means
out=work.gistemp_nh_month(rename=(col1=temp_anomaly _name_=month));
      var jan feb mar apr may jun jul aug sep oct nov dec; by year;
run;
data work.gistemp_nh_month;
      length date 8.;format date monyy5.;
      length type $50.; type="Northern Hemispheric Means";
      set gistemp_nh_month;
      date = input(compress(month || put(year,4.)), monyy7.);
      drop year month;
run;
proc append base=data.gistemp_monthly_anomaly data=work.gistemp_nh_month;
run;
proc transpose data=work.gistemp_nh_means
out=work.gistemp_nh_season(rename=(col1=temp_anomaly _name_=measure));
      var j_d d_n djf mam jja son; by year;
run;
data work.gistemp_nh_season;
      set work.gistemp_nh_season;
      length type $50.; type="Northern Hemispheric Means"; label
type="Region";
      label measure="Measure";
      label temp_anomaly="Temperature Anomaly (C)";
run;
proc append base=data.gistemp_season_anomaly data=work.gistemp_nh_season;
run;

/*  GISTEMP - Southern Hemispheric Means */
proc import datafile="&root.\gistemp\SH.Ts+dSST.csv"
out=work.gistemp_sh_means dbms=csv replace;getnames=yes;run;
proc transpose data=work.gistemp_sh_means
out=work.gistemp_sh_month(rename=(col1=temp_anomaly _name_=month));
      var jan feb mar apr may jun jul aug sep oct nov dec; by year;
run;
data work.gistemp_sh_month;
      length date 8.;format date monyy5.;
      length type $50.; type="Southern Hemispheric Means";
      set gistemp_sh_month;
      date = input(compress(month || put(year,4.)), monyy7.);
      drop year month;
run;
proc append base=data.gistemp_monthly_anomaly data=work.gistemp_sh_month;
run;

proc transpose data=work.gistemp_sh_means
out=work.gistemp_sh_season(rename=(col1=temp_anomaly _name_=measure));
      var j_d d_n djf mam jja son; by year;
run;
data work.gistemp_sh_season;
```

```
      set work.gistemp_sh_season;
      length type $50.; type="Southern Hemispheric Means"; label
type="Region";
      label measure="Measure";
      label temp_anomaly="Temperature Anomaly (C)";
run;
proc append base=data.gistemp_season_anomaly data=work.gistemp_sh_season;
run;

/*  GISTEMP - Zonal annual means */
proc import datafile="&root.\gistemp\ZonAnn.Ts+dSST.csv"
out=work.gistemp_zones_means dbms=csv replace;
      getnames=yes;
run;
proc transpose data=work.gistemp_zones_means
out=work.gistemp_zones_means(rename=(col1=temp_anomaly _name_=measure));
      by year;
run;
data data.gistemp_zones_means;
      length measure $25.;
      set work.gistemp_zones_means;
      label measure="Measure";
      label temp_anomaly="Temperature Anomaly (C)";
      measure = trim(left(tranwrd(measure, "_", " ")));
      if measure eq "Glob" then measure="Global";
      if measure eq "NHem" then measure="Northern Hemisphere";
      if measure eq "SHem" then measure="Southern Hemisphere";
run;
```

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Falko Shultz
SAS Institute Inc.
GPO Box 976
Brisbane, QLD 4001
Falko.Schulz@sas.com
http://www.sas.com