

CGT270

Midterm Part II

Data Visualization Challenge

Student:

Chia-Hua Lin

Professor

Dr. Vetria Byrd

TA

Subia Ansari

Course:

CGT 27000 Data Visualization

Term:

Fall 2021

Contents

Halloween Visualization	3
Data Description	3
Location of home	4
Example	4
The Assignment	5
Data Visualization Process	6
Acquire	6
The Data	6
Parse & Mine	7
Represent	8
Filter	10
Critique	11
Refine	12
What's the story?	14

Halloween Visualization

This in-class assignment is to create data visualizations using data collected about trick-or-treaters in Cincinnati, OH. [You should create two \(2\) visualizations](#), this can be a collection of charts or a dashboard, whatever is necessary to the story or analysis that is shown in your visualizations. Make sure you [READ and FOLLOW ALL Instructions](#). The goal is to demonstrate your understanding of the data visualization process.

Data Description

The data is available in two formats

- Halloween data for Excel 2020 is a crosstab table which is ideal for creating visualizations in Excel. Numbers in the data file for Excel are **cumulative**.
- Halloween data for Tableau 2020" is unpivoted which is ideal for creating visualizations in Tableau. Numbers in the data file for Tableau are **not cumulative**.
- The data has been collected since 2008.
- The numbers in the table are cumulative totals of the number of trick-or-treaters who visited one house each year.
- The numbers are measured at 30-minute intervals, except for the last 15-minute interval.
- The trick-or-treat count was recorded in 30-minute intervals except for the last 15-minute interval.
- The night of trick-or-treating has always been on October 31st each year (some neighborhoods change the night of trick-or-treating).
- Official tick or treat hours are from 6 PM to 8 PM, but there are often "stragglers" past 8 PM that are not turned away. These stragglers are counted in the 8PM – 8:15 PM time slot. There has never been a trick-or-treater past 8:15 PM.
- The type of candy did not vary year-by-year. It is always a general mix of candy purchased in bulk variety bags.

Location of home

Neighborhood: East Walnut Hills/Evanston

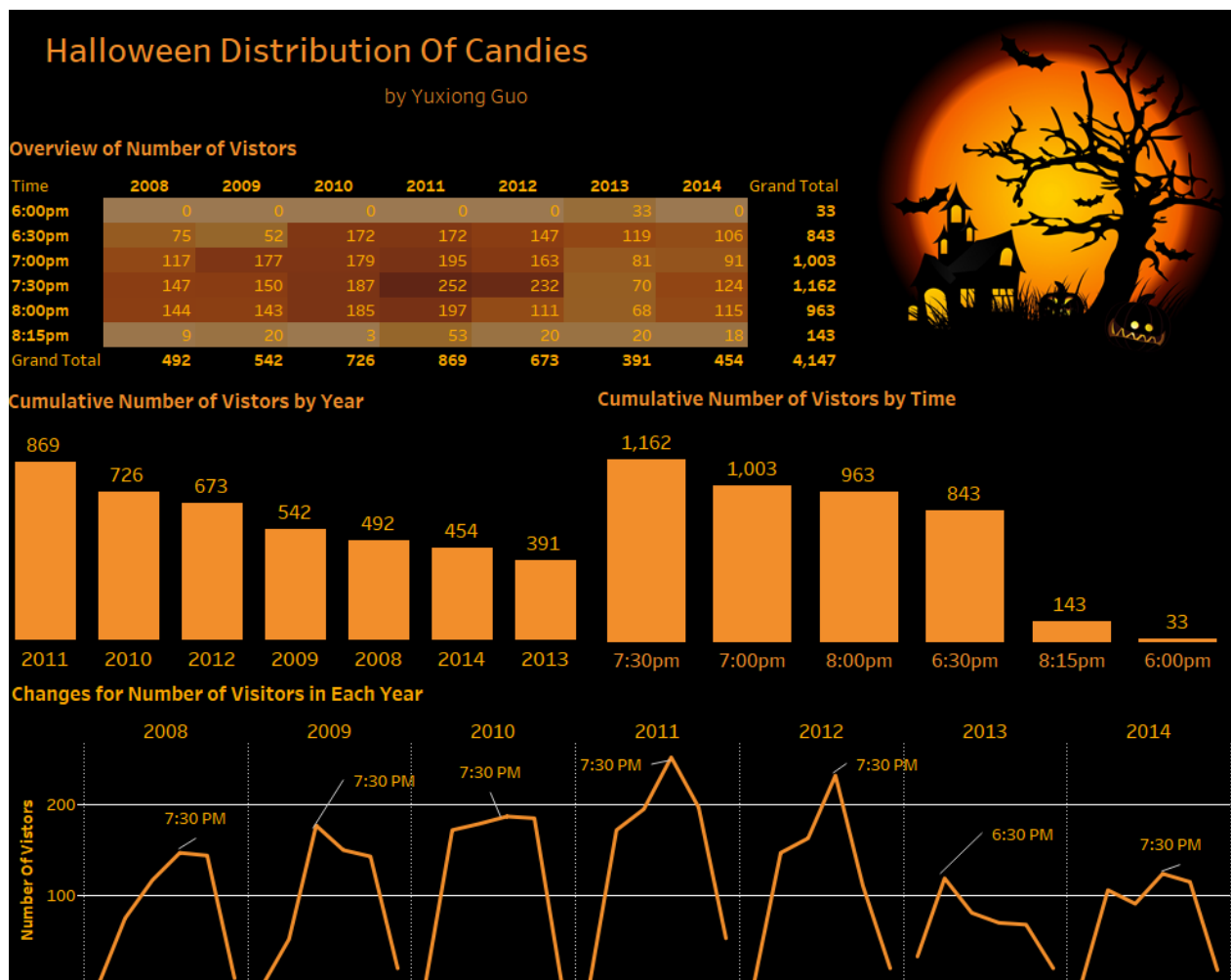
City, State: Cincinnati, Ohio

Zip code: 45207

Being a corner house on the neighborhood border likely increases the number of trick-or-treaters.

Example

Here' an example of how previous Halloween data have been visualized. Be creative!



The Assignment

There are multiple parts to this assignment. Make sure you read the entire assignment before starting.

Determine a story or goal to support the two (2) visualizations you will create using the Halloween data provided. Your two visualization **MUST** be different chart types. **This means DO NOT create two bar charts or two-line charts or two of the same chart types!** Challenge yourself. This is your time to show what you know.

Examples (these are examples):

- Homeowner dashboard summarizing Halloween
- Forecast future trick-or-treaters or estimate future candy needed
- Explore variation of the number of trick-or-treaters year by year
- [Be creative and think of other things you could do](#)

Data Visualization Process

Show your understanding of the data visualization process.

Acquire

The Data

Year	6pm	6:30pm	7pm	7:30pm	8pm	Total (8:15pm)
2020	11	55	107	155	211	219
2019	0	117	262	406	483	523
2018	18	191	342	497	589	600
2017	41	190	357	549	710	776
2016	22	160	386	612	759	822
2015	13	148	336	523	667	747
2014	0	106	197	321	436	454
2013	33	152	233	303	371	391
2012	0	147	310	542	653	673
2011	0	172	367	619	816	869
2010	0	172	351	538	723	726
2009	0	52	229	379	522	542
2008	0	75	192	339	483	492

Excel and Tableau versions of the data are provided in Brightspace. Choose one (1) to work with.

- HalloweenExcel
- HalloweenTableau

Parse & Mine

Use this page to provide a parsing of the data. For quantitative fields list some basic statistical procedures that can be performed in the space below. To be clear, you are to list the procedure (you are not required to actually do any calculations here).

Use the Tab key to add more rows to the table below.

Variable	Data type	Statistical Method (where applicable)
Date	String	
Date and Time	String	
Count	Integer	Average, Minimum/Maximum, Standard deviation
Day of Week	String	
Time	String	

Represent

How to replace this figure: Right-click on the figure below, select Change Picture → From a File. Locate your figure.

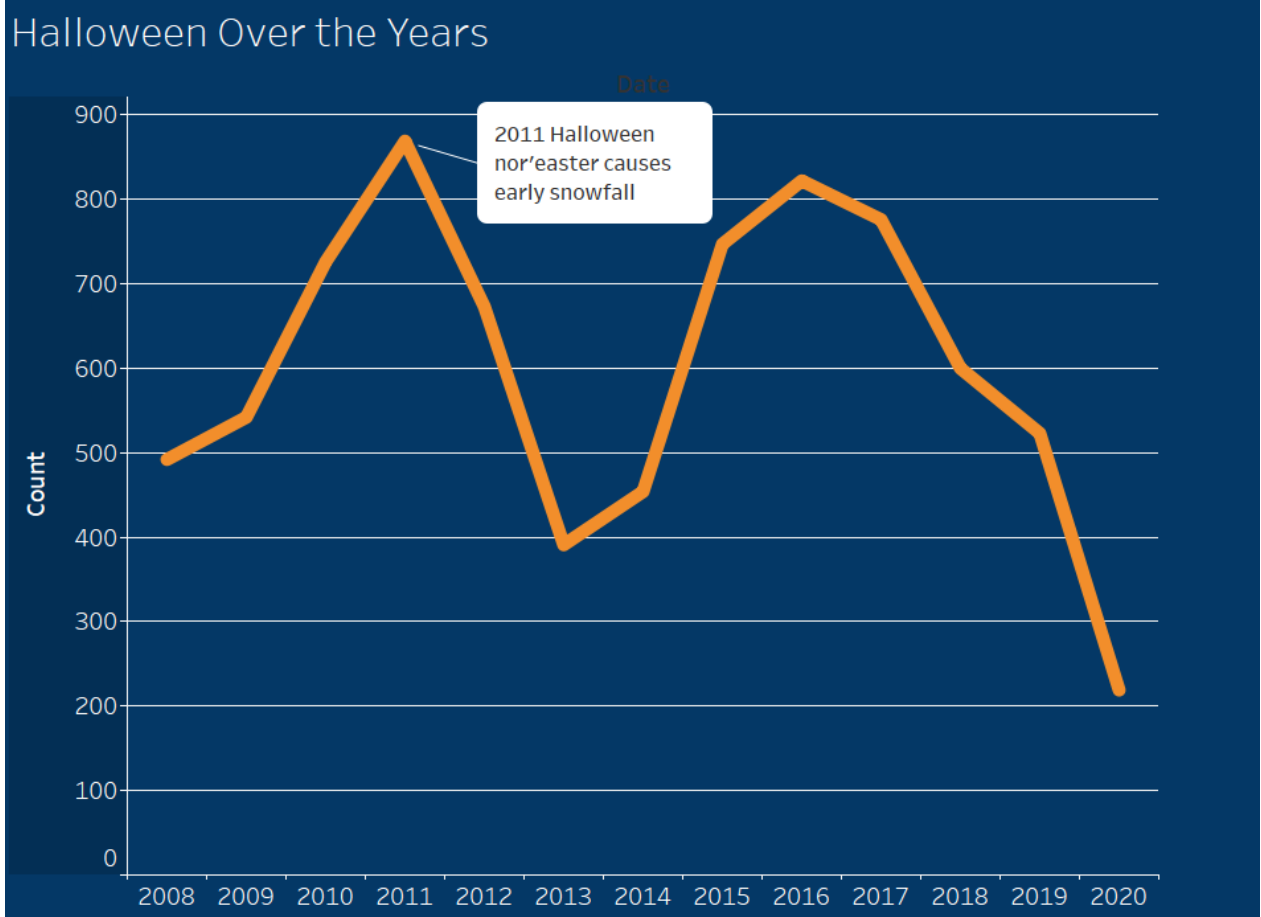
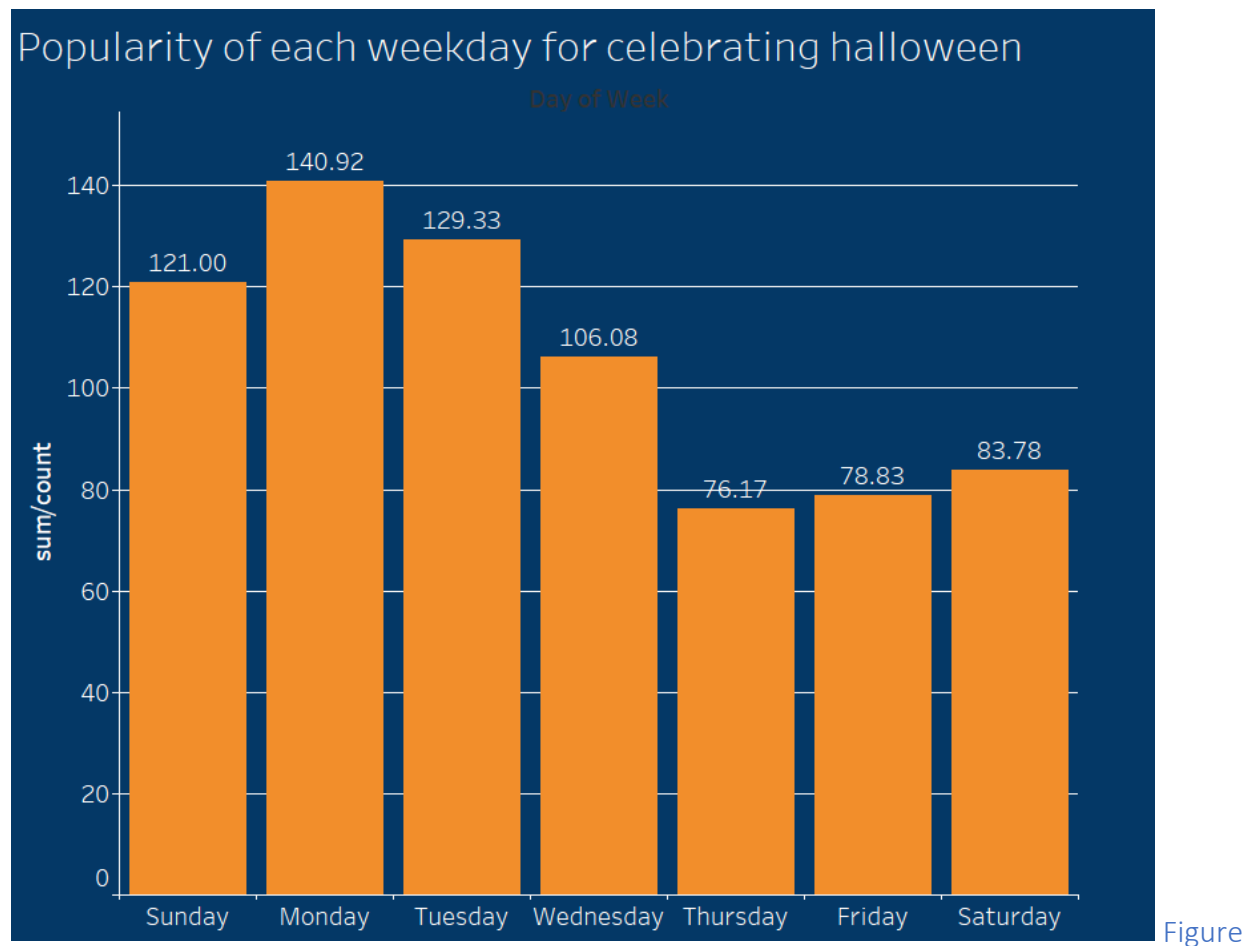


Figure 1. Line graph depicting the count of people participating in Halloween over the years.
Source: Chia-Hua Lin

How to replace this figure: Right-click on the figure below, select Change Picture → From a File. Locate your figure.



2. Bar graph depicting the popularity of each weekday as Halloween celebration using the average visitors each half hour. Source: Chia-Hua Lin

Helpful Tip: Utilize the space that you have. Do NOT create a tiny visualization that is unreadable. Remember, the purpose of visualization is insight, but all insight is lost if it cannot be seen.

Use this page format for visualization that require a landscape layout. Remove this text and replace the figure with your own visualization.

Filter

In this page show the data you used to create your visualizations.

Figure 1

I used the Date and the Count data fields to create my visualization for Figure 1.

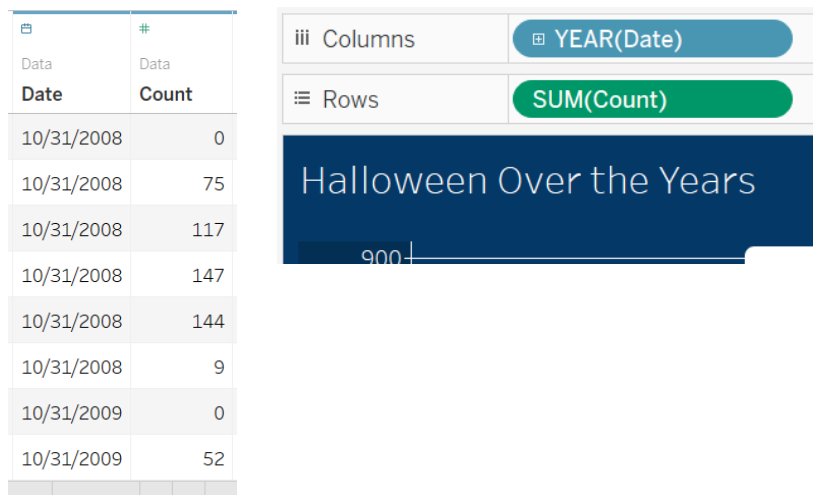
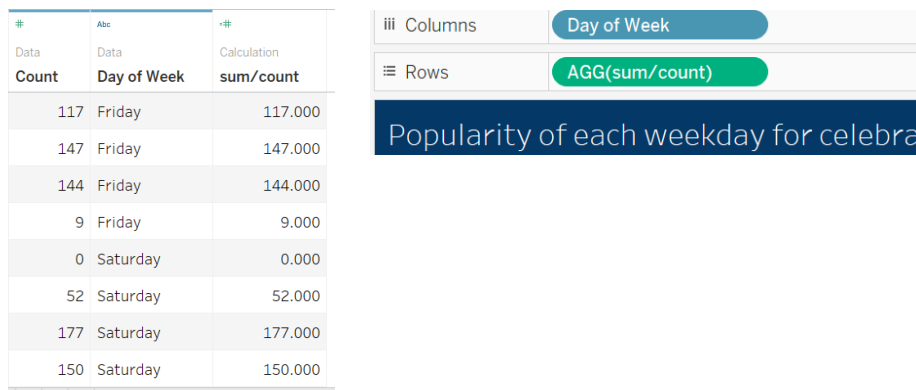


Figure 2

I used the Count and Day of Week data fields, as well as a separate calculated field which I named sum/count and aggregated to create my visualization for Figure 2.



Critique

Rate your visualizations (Figure 1 and Figure 2) using the link below

<https://stephanieevergreen.com/rate-your-visualization/>

Figure 1 Rating

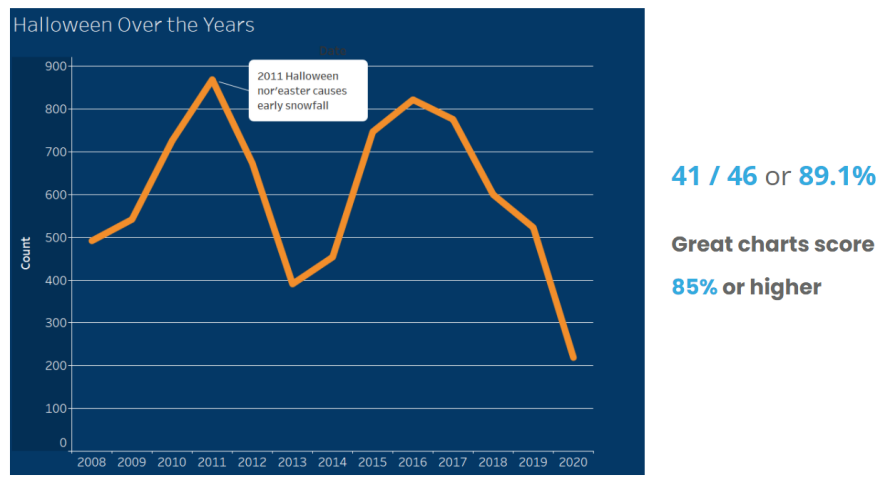
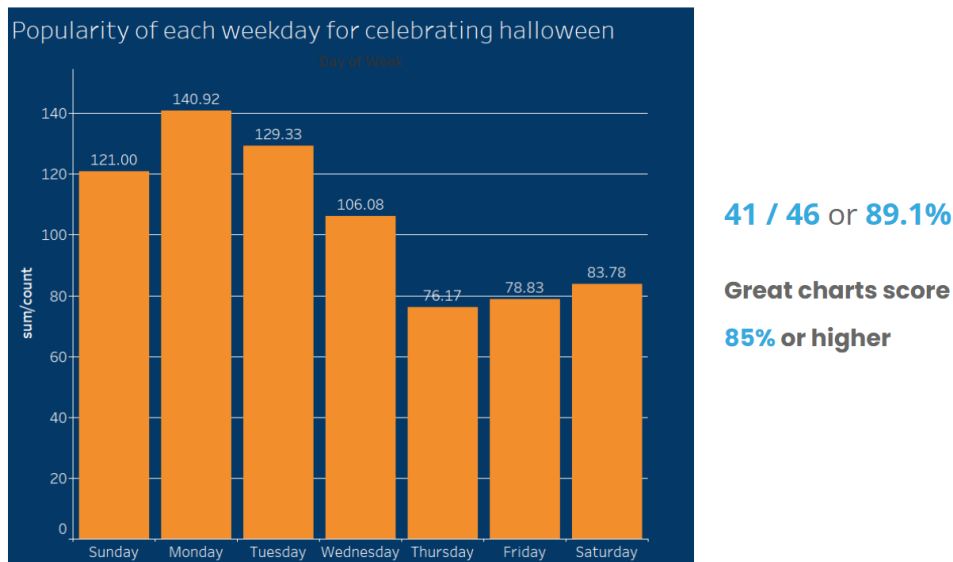


Figure 2 Rating



Refine

In this part of the visualization challenge, you should identify one or more characteristics of the visualizations you created (Figure 1 and Figure 2) and update the figures. Include an updated version of each Figure below. In the figure caption, state what changes were made.

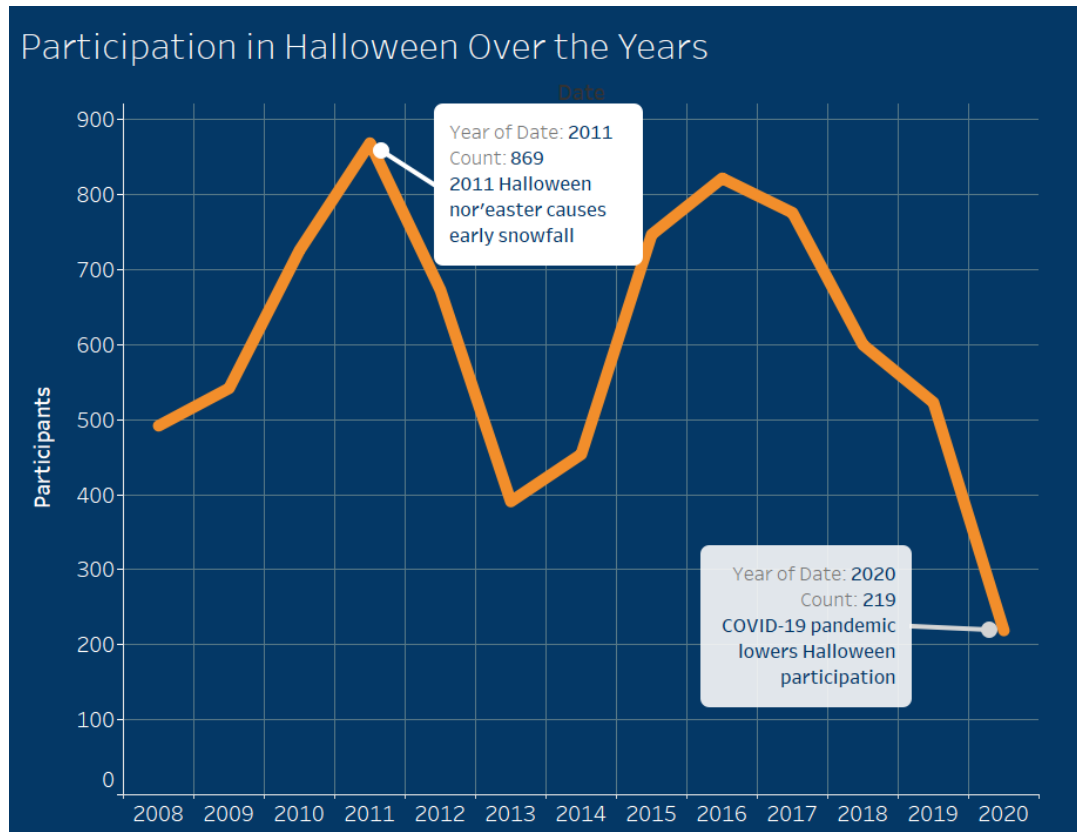


Figure 1 Refined. Changed title to be more concise. Updated background color of Y-axis labels and annotation font for consistency. Renamed Y-axis title from 'Count' to 'Participants'. Added annotation for 2020 event. Changed grid lines from white to dark blue. Added vertical grid lines for clarity. Source: Chia-Hua Lin

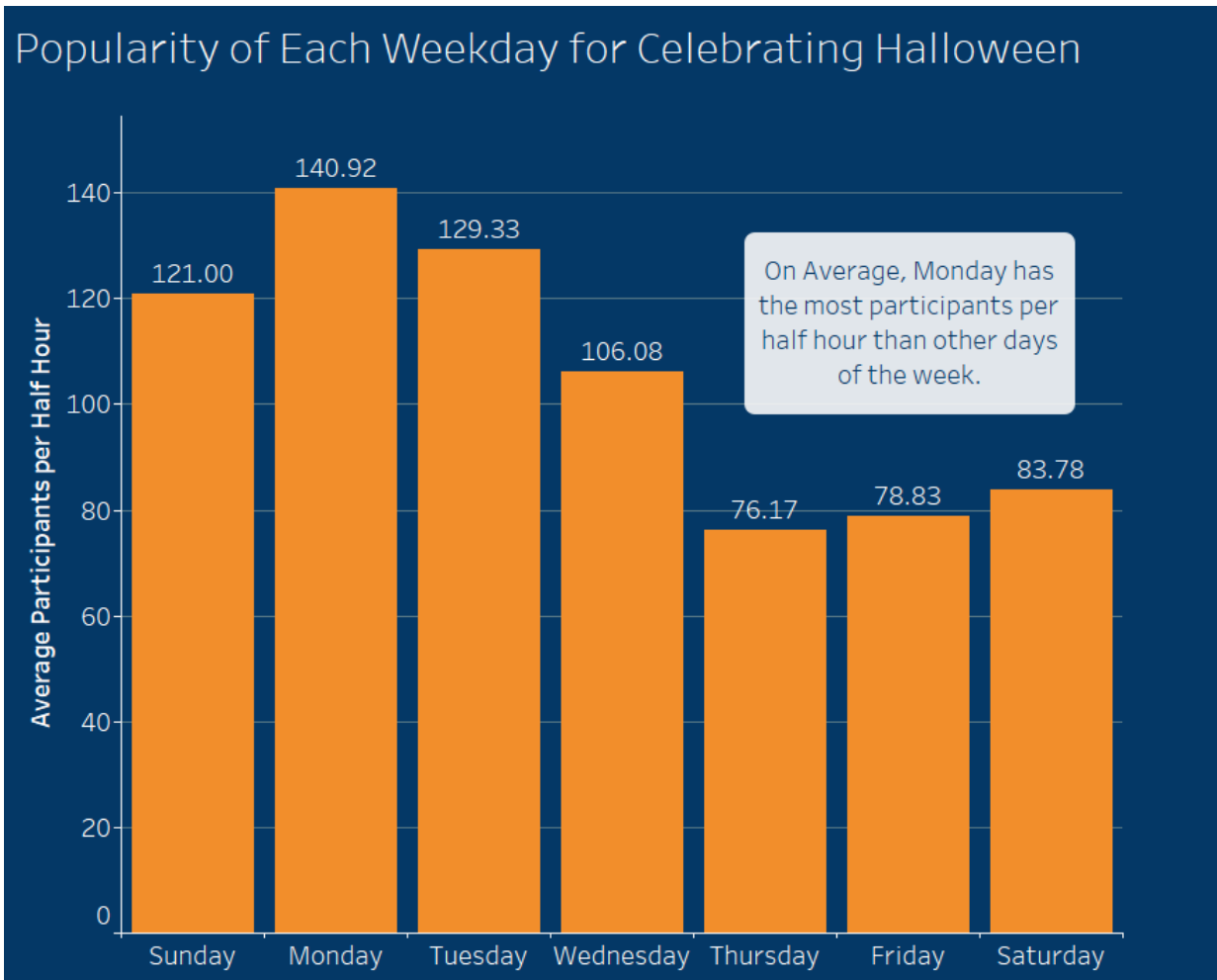


Figure 2 Refined. Deleted stray text. Changed Y-axis title. Deleted vertical grid lines, added muted horizontal grid lines. Added annotation. Source: Chia-Hua Lin

What's the story?

The story I told using the data was about how Halloween has changed over time. In Figure 1, we can see that events such as the Halloween nor'easter and COVID-19 have drastically affected people's participation in Halloween. The Halloween nor'easter was a large low-pressure area, which caused it to snow very hard from October 28th to November 1st of 2011, which was why Halloween participation increased. COVID-19 in the year 2020 made it so that there were less people going out to participate in Halloween, because people wanted to socially distance themselves and prevent the spread of the disease as best as they could. There were also variations in the popularity of Halloween throughout the week, as seen in Figure 2; Monday is the most popular day for participation, and Thursday is the least popular.

Besides this midterm, my data visualizations could potentially be used by a marketing team to observe the general trend of popularity and its relationship with time for Halloween.

One assumption I made about the data was that the Count field measured the amount of trick-or-treaters' participation in Halloween. Another assumption was that the home collecting the data recorded every person regardless of age or costume. I also assumed that there were no people' participation in Halloween before 6:00pm or after 8:15 pm, meaning that the data we have right now is complete.

Checklist of what to submit:

- Save this file as LastnameFirstInitial_CGT270Fall2021_MidtermPartII.pdf
- Only submit one (1) file. All of your work should be contained in this file.
- Failure to follow these instructions will result in your work NOT being graded.

General Deductions (others made accordingly)

- No name on the first page of the document: -5 pts
- Altered template: -10 pts
- No figures included: -15 pts for each missing figure
- No figure captions: -10 pts for each missing caption
- Zip file submitted: See Checklist of what to submit (-80 pts)
- Late submissions: Will NOT be graded (-80 pts)
- Provided a link to visualizations instead of providing screenshot of the visualization: this will be treated as no figure, no figure caption (-25 pts)
- Failure to follow data visualization best practices (data visualization checklist): deductions made appropriately.

Keep in mind: one (1) second after the submission
deadline is considered late.



Byrd Data Visualization Lab