

# CGT270

# Midterm Part II

## Data Visualization Challenge

Student:

Thomas Cluff

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.....	Error! Bookmark not defined.
Filter .....	9
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

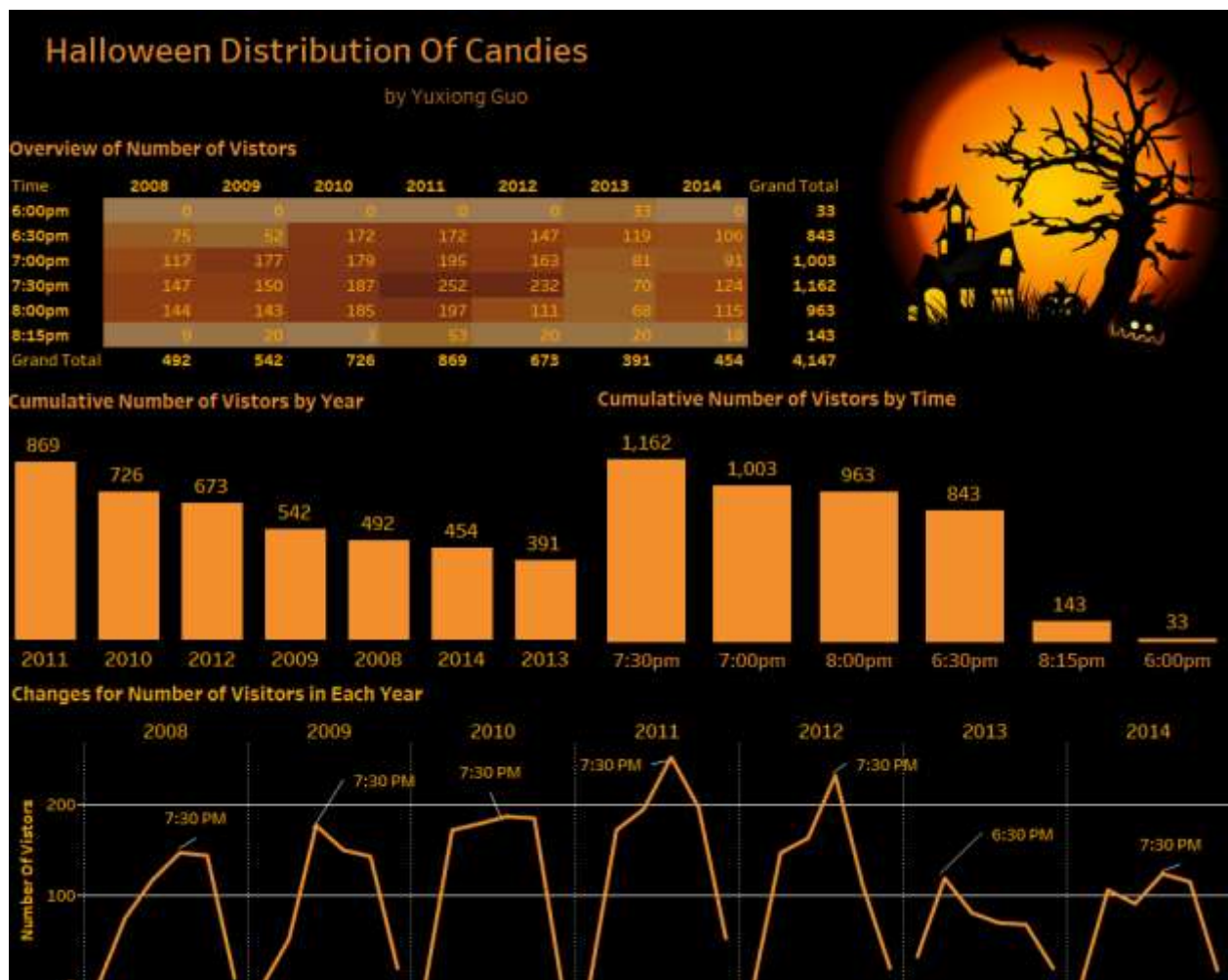
City, State: Cincinnati, Ohio

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	Date data	You can calculate the mode, median, mean, max, and min.
Date and Time	Date and Time data	You can calculate the mode, max, and min.
Count	Integer data	Calculate the median, mode, mean, max, min, and standard deviation
Day of the Week	String data	You can find the mode, and order the days
Time	String Data	If you convert to a number you can find the median, mean, mode, max, min, and standard dev. But keeping it as a string only the mode, median, mean, max, and min can be found.

# Represent

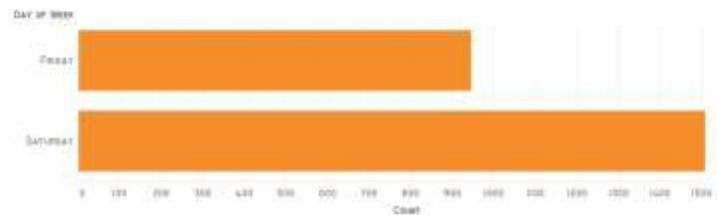
## HOW SCHOOL NIGHTS AFFECT TRICK OR TREATER COUNT

### WEEKNIGHTS



As you can see there is no consistency between counts on weekends or school nights. However if you look at the count over the years, there is a clear smooth relation between the years. So trick or treater counts depends more on the yearly social climate than on what day of the week Oct 31st is on that year.

### WEEKENDS



### COUNT THROUGH THE YEARS

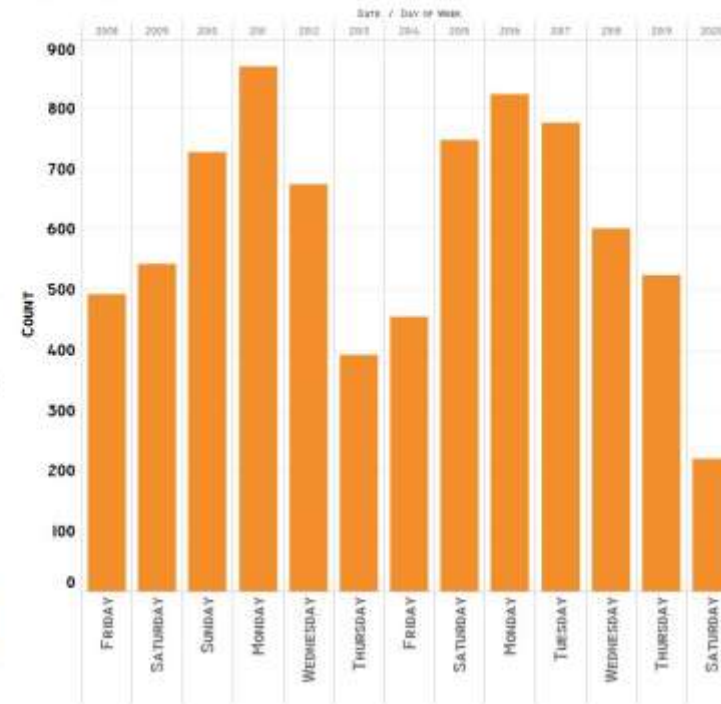


Figure 1: As you can see there is no consistency between counts on weekends or school nights. However, if you look at the count over the years, there is a clear smooth relation between the years. So, trick or treater counts depends more on the yearly social climate than on what day of the week Oct 31st is on that year.



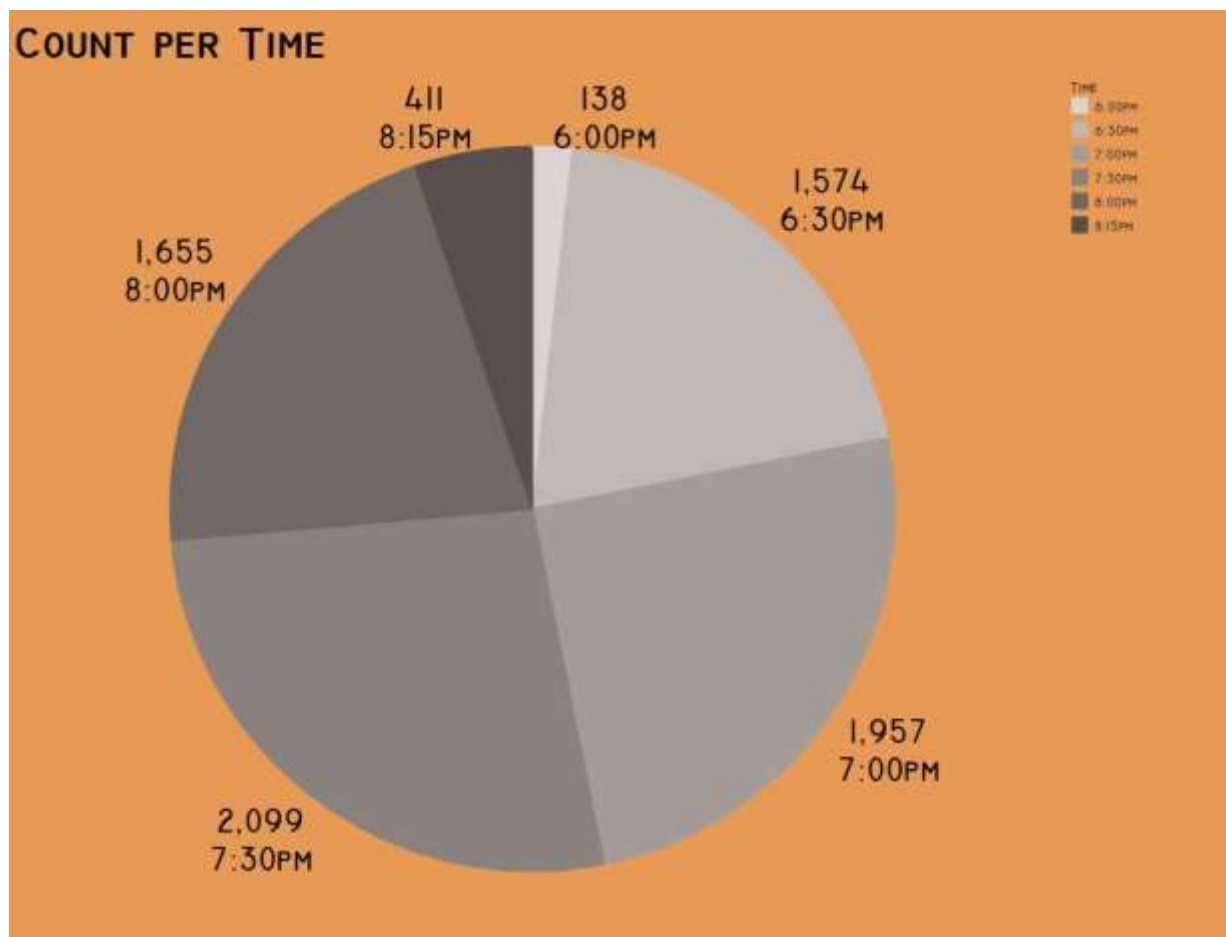


Figure 2. This image shows how the count of trick or treaters is divided amongst the time slots measured

**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.

# Filter

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

## Figure 1

For figure one, I used the day of the week, the year, and the number of trick or treaters. I filtered the count of trick or treaters by what day Halloween occurred and also by what year. Then in the visualization I compare the two trends

## Figure 2

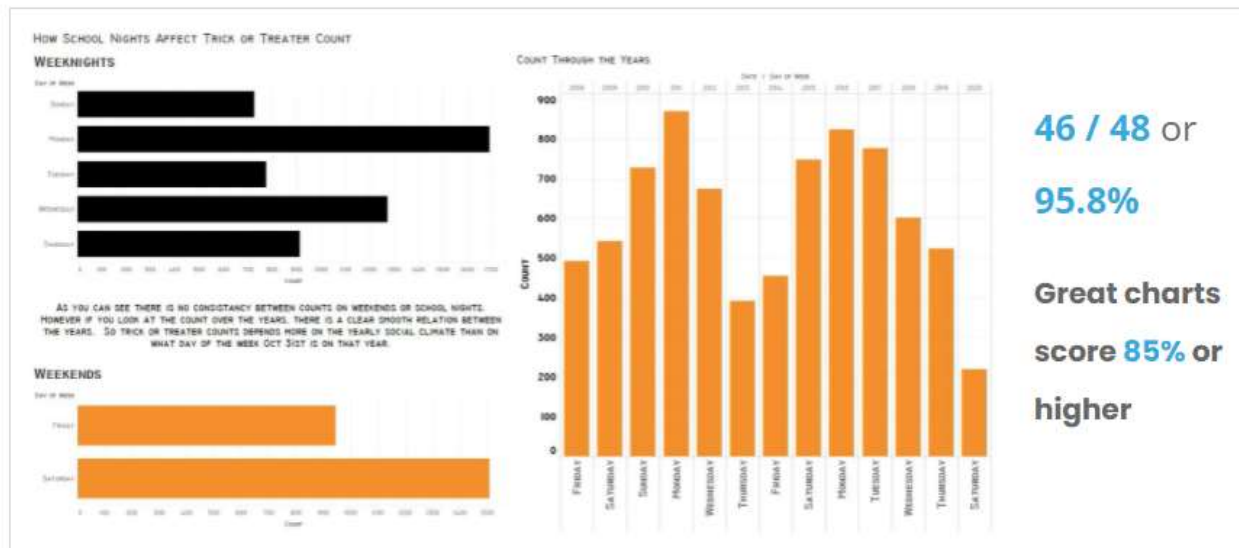
For figure two, I used the time data and the count of trick or treaters. I created a pie chart to show what portion of trick or treaters like to go out at what time.

# Critique

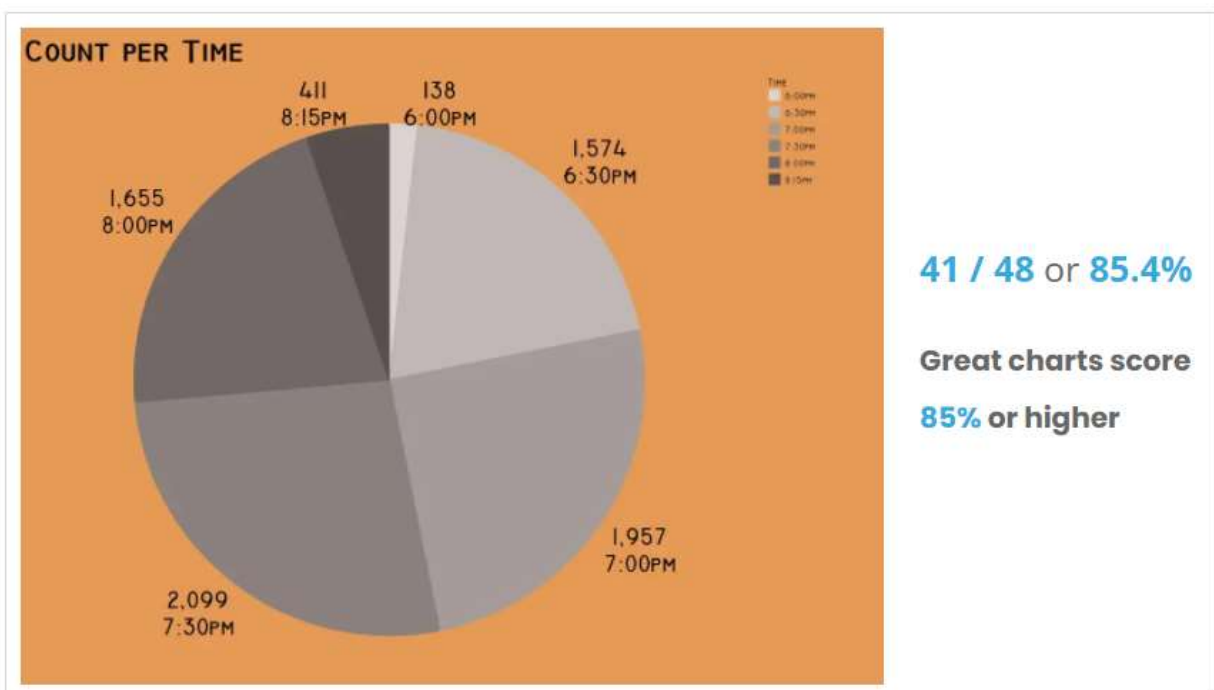
Rate your visualizations (Figure 1 and Figure 2) using the link below

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

## Figure 1 Rating



## Figure 2 Rating



# Refine

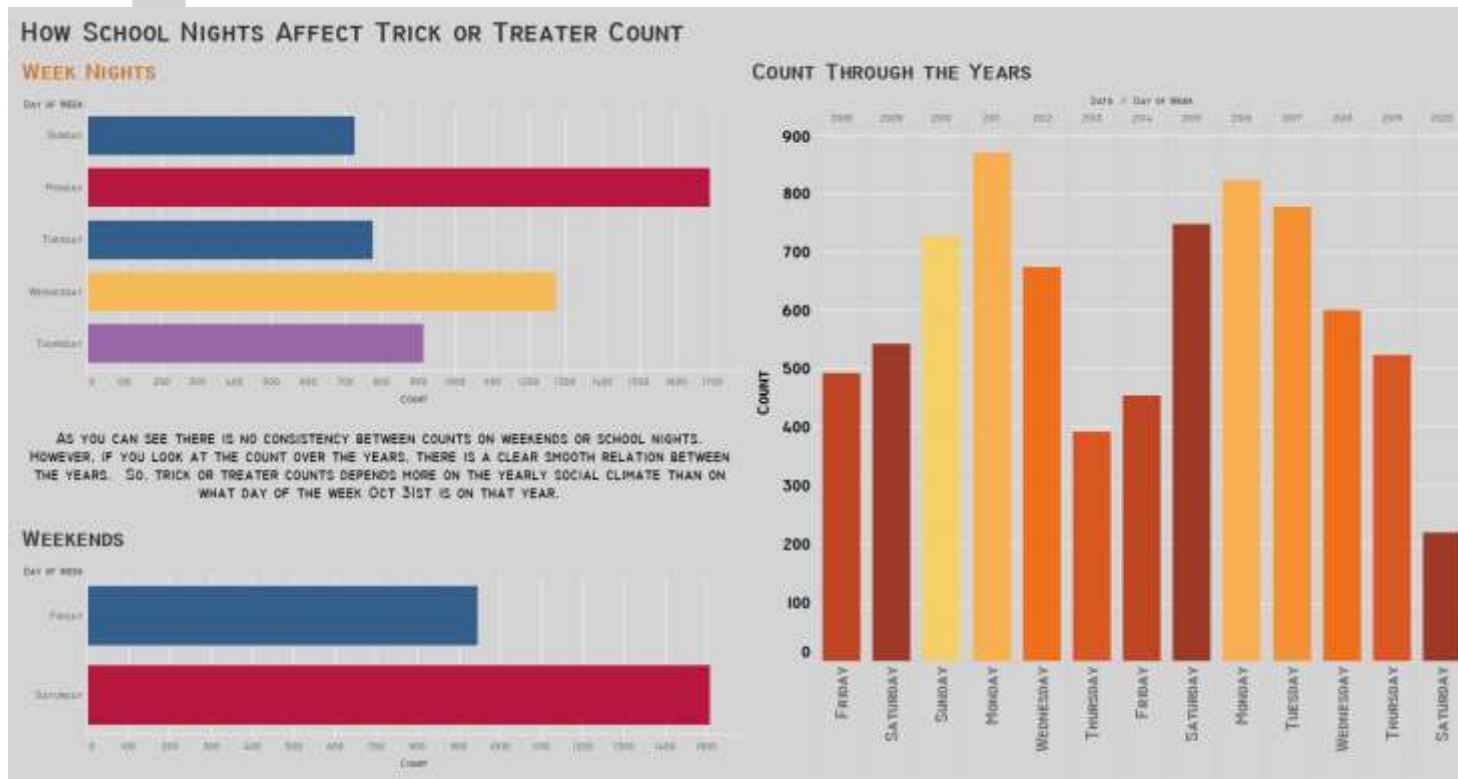


Figure 1 Refined: To refine this visualization I changed up the color schemes to be a little more dynamic, I changed the background shading, and I adjusted the text sizing to fit a better hierarchy. The color decisions were themed around sunsets and fall because that is when children trick or treat

## TIME AVERAGES THROUGHOUT THE WEEK

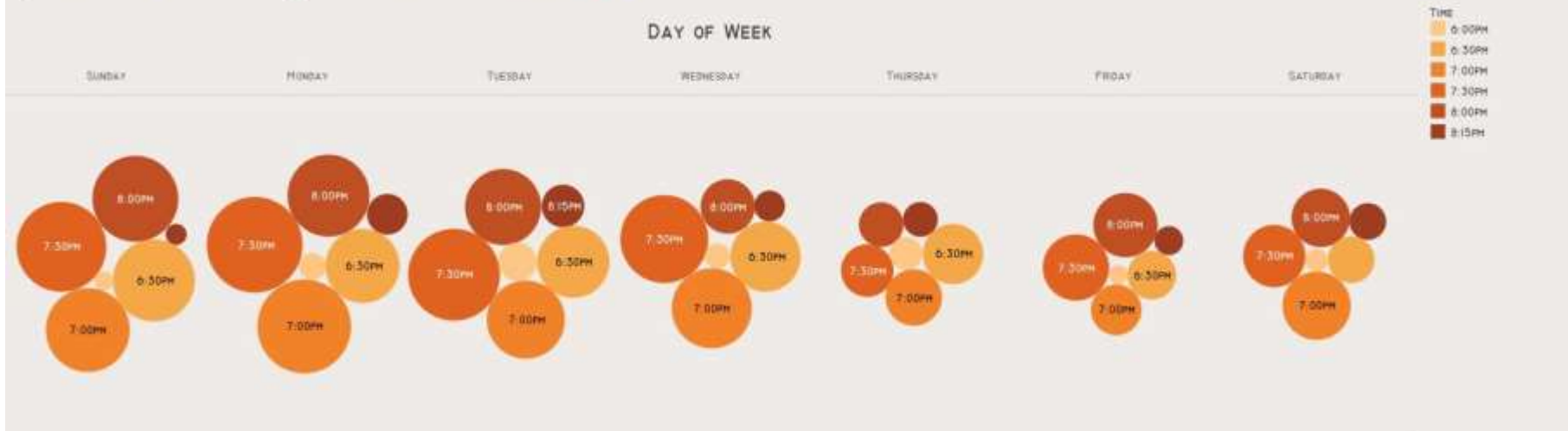


Figure 2 Refined: When filling out the critique I realized that my whole question did not suit how I was visualizing my data. So, I changed my question to “How does day of the week affect when trick or treaters go out”. By switching from a pie chart to a cluster and breaking that cluster up depending on the day, I am able to show what times are most common for that day of the week. I also adjusted the color scheme because it was not easy to look at.

# What's the story?

After looking through the data provided I wanted to cover when children trick or treat because I felt like the data had a lot of detail for date and time and also had an emphasis on how many kids were trick or treating. For Figure one I began by seeing if there was a correlation between oct 31<sup>st</sup> being on a school night with the amount of trick or treaters. When visualizing the data, I found that there was no consistency. I then decided to graph trick or treaters by year and found that there were smooth transitions between years. This led my first figure to show that the amount of trick or treaters in a year is dependent on the social climate that year and not on what day of the week Halloween is on.

For my second figure, I wanted to cover what time children would trick or treat depending on the day of the week, rather than year. By graphing this out using cluster charts, I found that during the first half of the week, there were a lot more trick or treaters. However relative to each time slot, the amount of trick or treaters changes. The closer the time slot is to the middle (so closest to 7:00p, - 7:30pm), the bigger the drop in trick or treaters later in the week. For children going out at 6:00pm there is almost no difference in the day of the week.

My audience for these visualizations is homeowners who want to know how much candy to buy for a given year, and when to be most actively waiting by the door. If you know how many kids will come depending on the week day, you can budget your candy expenses better. If you know when they will arrive, you can plan your day around when the doorbell will be rung the most.

I made three assumptions when analyzing the data. My first assumption for this data is that more children would trick or treat on non-school nights. But this was quickly proven wrong when visualizing. My second assumption is that this data is consistent across all neighborhoods with a similar demographic to East Walnut Hills/Evanston in Ohio. The exact numbers may not match but the relative proportions should. My third assumption is that numbers dropped in 2020 due to the Covid-19 pandemic, so the decline in count for that year is explained and does not need to be addressed.

**Bonus points for REALLY GOOD stories!**

### 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