

## Information Visualization - Homework/Discussion 2

**Tufte Principles. Explore a hypothesis in  
E. Repens incidence.**

---

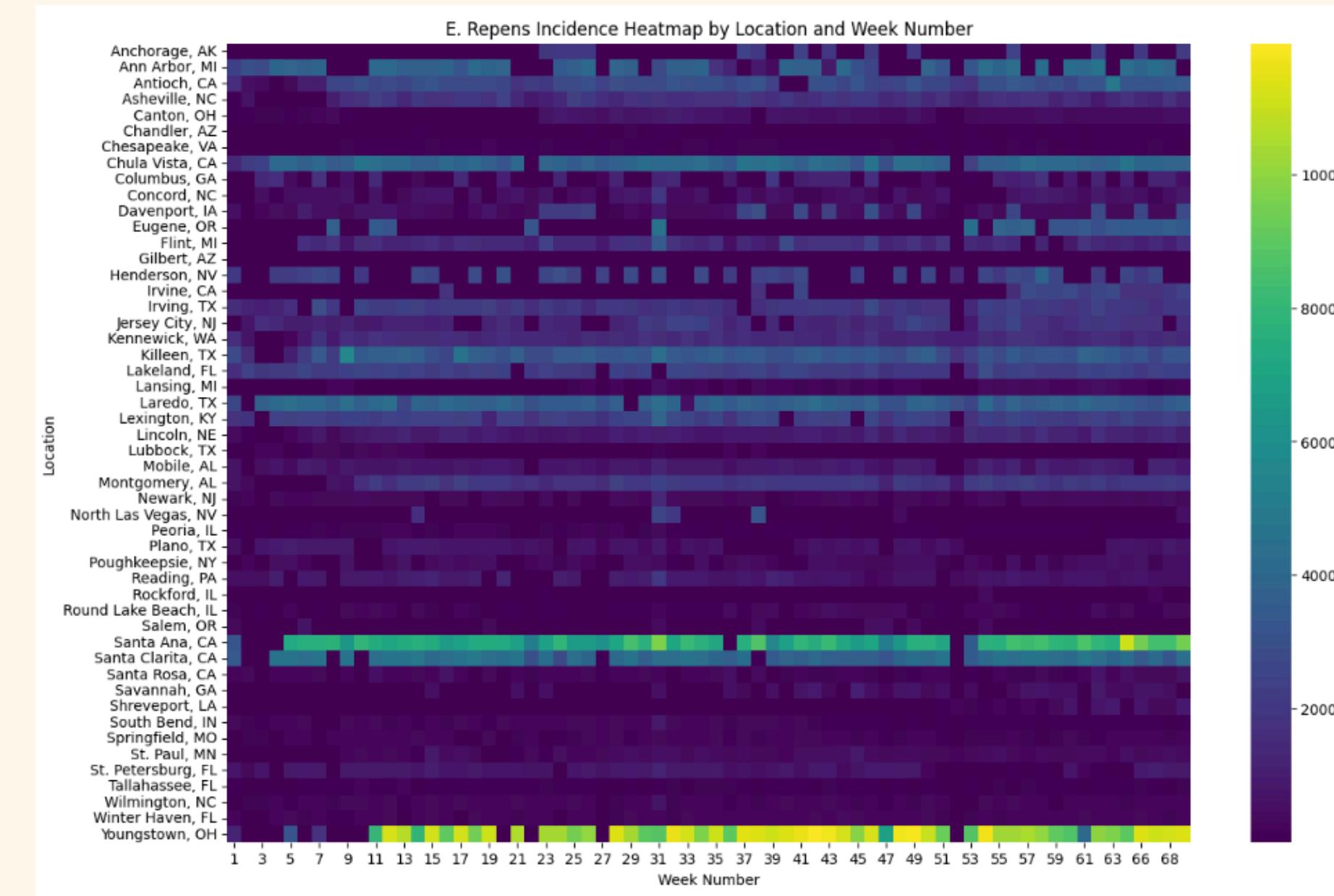
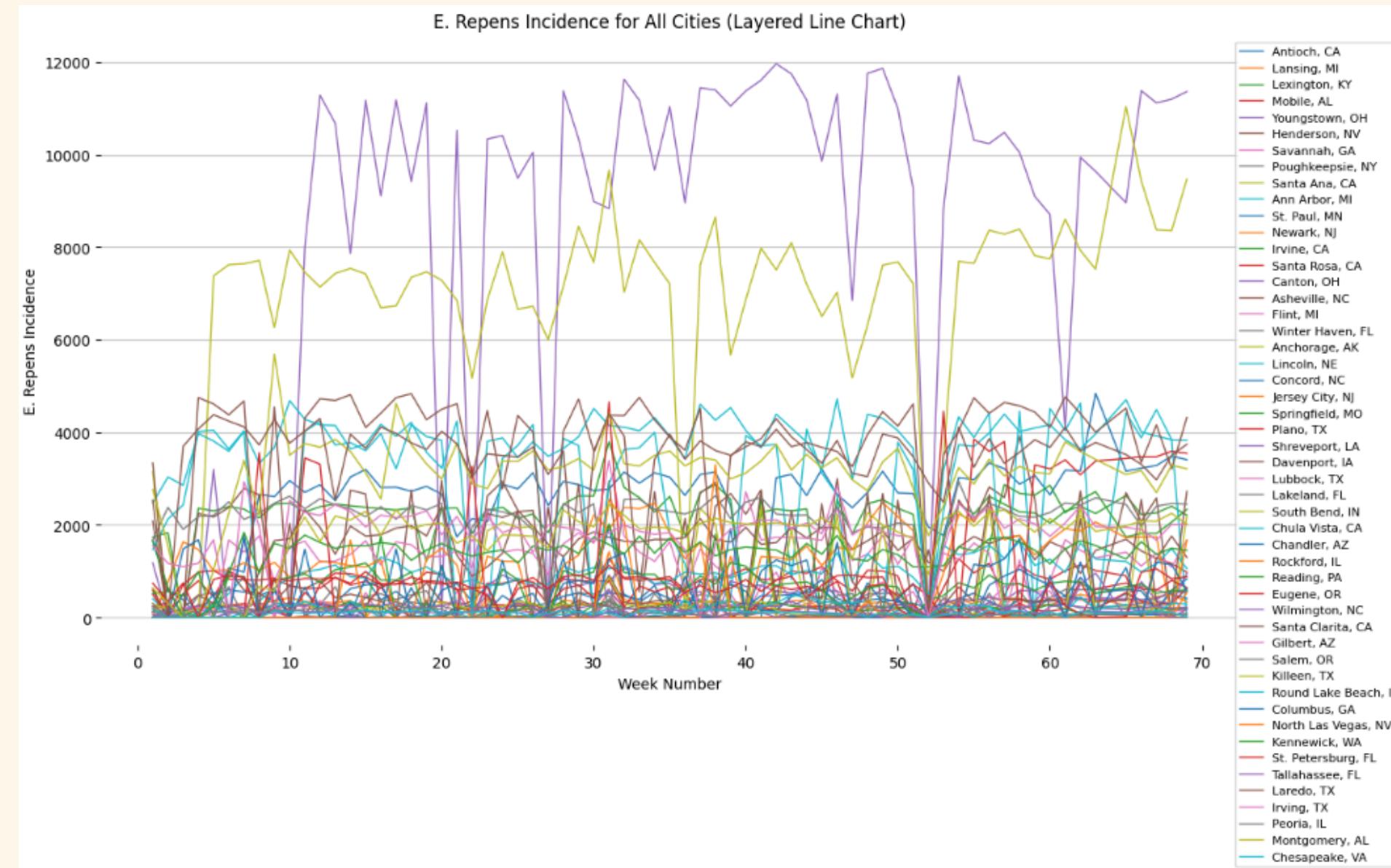
## Purpose

This analysis explores the E. Repens incidence numbers for US cities over consecutive weeks, as provided in the "ERepens.csv" dataset. The primary goal is to investigate whether there is a periodicity in the incidence data for different cities, specifically looking for patterns with periods of 6, 10, 14, and 18 weeks. The analysis aims to identify cities that exhibit similar periodic trends and visualize these patterns to aid in understanding the data and potentially grouping cities.

## Dataset - **ERepens.csv**

The dataset in the ERepens.csv file contains E. Repens incidence numbers for various US cities over a number of consecutive weeks. It includes columns for the location (city and state), the reporting year, the week number within that year, and the corresponding E. Repens incidence value. The dataset has 5 columns and 3400 rows.

# Times Series plots for each location overlapped in a line chart and HeatMap of incidents to visually inspect for periodic



These charts allow for visual inspection of recurring patterns and comparison of trends across cities within each group.

By examining this chart, you can:

- Observe the overall trend of E. Repens incidence for each city over the entire period.
- Compare the incidence levels and patterns of different cities directly on the same axes. You can see if cities experience peaks or valleys at similar times or if their trends differ significantly.
- Visually look for any recurring patterns or cycles within individual city lines, although with many lines on the same plot, this can sometimes be challenging to see clearly for every city.

By looking at the patterns of color across the rows (for a single city over time) and down the columns (for all cities in a specific week), you can visually identify:

- Weeks or periods with generally high or low incidence across many cities.
- Cities that consistently have higher or lower incidence than others.
- Potential cyclical patterns or spikes in incidence for individual cities or groups of cities as color intensity changes over the weeks.

## How Tufte's principle was applied

Line charts were used to visualize the E. Repens incidence trends over time for cities grouped by potential periodicities (6, 10, 14, and 18 weeks).

**Adherence to Tufte's Principles:** The visualizations were designed to maximize the data-ink ratio and minimize chartjunk by: Removing unnecessary chart elements like heavy borders and excessive grid lines

Focusing on the data itself (the lines representing incidence over time) and ensuring clarity and precision.

These charts allow for visual inspection of recurring patterns and comparison of trends across cities within each group.

## Findings in Periodicity Visualizations adhering to Tufte's principle

### 6-week Periodicity Group:

**Cities:** 'Lansing, MI', 'Savannah, GA', 'Irvine, CA', 'Canton, OH', 'Shreveport, LA', 'South Bend, IN', 'Eugene, OR', 'St. Petersburg, FL', 'Peoria, IL'

**Observation:** The line charts for this group generally show recurring peaks in E. Repens incidence approximately every 6 weeks. The amplitude and sharpness of these peaks vary between cities, suggesting differences in the intensity and timing of outbreaks within this cyclical pattern.

### 10-week Periodicity Group:

**Cities:** 'South Bend, IN'

**Observation:** Only South Bend, IN was in this group. The line chart for South Bend shows fluctuations, and while there are some recurring increases, a clear and consistent 10-week periodic pattern in incidence peaks is not strongly visually evident across the entire time series compared to the 6-week patterns observed in the other group. Shorter periodicities also appear to be present.

### 14-week and 18-Week Periodicity Group:

**Observation:** No cities were identified with a significant 14-week and 18-Week periodicity in the initial analysis.

