Lyme Disease

James Park and Nikolai Rabin
Multimedia, Hypertext and Information Access
CS 4624
Edward A. Fox
Blacksburg, VA 20461
May 3, 2021
Outline

1. Project Background and Requirements
2. Software Used
3. Design Approach
4. Web Scraping
5. Generation of First Plots
6. Exploration of Plots and Best Predictors
7. Choropleth Graphs and Final Findings
8. What We've Learned/Future Work
9. Acknowledgements
10. References
Project Background and Requirements

- **Project Background:**
  - What is Lyme Disease and why is it important?

- **Requirements:**
  - Given the number of Lyme disease cases per county, determine what could be causing these numbers.
  - Gather, plot, and analyze Lyme Disease data
Software Used

- RStudio
- Excel
- Google/Google Drive
Design Approach

1.) Web scrape data that could be potential predictors of Lyme Disease cases
2.) Clean the data that was scraped and remove unnecessary fields
3.) Explore the data that was scraped and create a series of plots for visualization
4.) Determine the best predictor variables
5.) Create choropleth graphs to visualize our findings
Web Scraping

- Utilize “rvest” library in R to scrape data from Wikipedia and other sources
- Data collected on:
  - Human Population Density by county
  - Human Population Count by county
  - Per Capita Income by county
  - Human Development Index by county
  - Temperature and Precipitation by county
  - Research and Development Spending per state
<table>
<thead>
<tr>
<th>Rank</th>
<th>State</th>
<th>Territory</th>
<th>HDI (2018) [note 1]</th>
<th>Comparable country [note 2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Massachusetts</td>
<td>Very high human development</td>
<td>0.956 Norway</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Connecticut</td>
<td>Very high human development</td>
<td>0.953 Norway</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Minnesota</td>
<td>Very high human development</td>
<td>0.947 Switzerland</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>New Jersey</td>
<td>Very high human development</td>
<td>0.943 Ireland</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>New Hampshire</td>
<td>Very high human development</td>
<td>0.943 Ireland</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>District of Columbia</td>
<td>Very high human development</td>
<td>0.942 Ireland</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Colorado</td>
<td>Very high human development</td>
<td>0.942 Ireland</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>New York</td>
<td>Very high human development</td>
<td>0.942 Ireland</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>North Dakota</td>
<td>Very high human development</td>
<td>0.942 Ireland</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Hawaii</td>
<td>Very high human development</td>
<td>0.94 Germany</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Washington</td>
<td>Very high human development</td>
<td>0.94 Germany</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Alaska</td>
<td>Very high human development</td>
<td>0.937 Sweden</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Vermont</td>
<td>Very high human development</td>
<td>0.936 Singapore</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Maryland</td>
<td>Very high human development</td>
<td>0.935 Singapore</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Nebraska</td>
<td>Very high human development</td>
<td>0.934 Singapore</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Wyoming</td>
<td>Very high human development</td>
<td>0.932 Netherlands</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Delaware</td>
<td>Very high human development</td>
<td>0.932 Denmark</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>California</td>
<td>Very high human development</td>
<td>0.932 Denmark</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Utah</td>
<td>Very high human development</td>
<td>0.932 Denmark</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Oregon</td>
<td>Very high human development</td>
<td>0.929 Denmark</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Illinois</td>
<td>Very high human development</td>
<td>0.929 Denmark</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Virginia</td>
<td>Very high human development</td>
<td>0.929 Denmark</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>South Dakota</td>
<td>Very high human development</td>
<td>0.929 Denmark</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Wisconsin</td>
<td>Very high human development</td>
<td>0.929 Finland</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Iowa</td>
<td>Very high human development</td>
<td>0.929 Finland</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Minnesota</td>
<td>Very high human development</td>
<td>0.929 Finland</td>
<td></td>
</tr>
</tbody>
</table>

Note 1: HDI is the Human Development Index.
Note 2: Comparable countries are listed for reference.
Generation of First Plots

- **Statewide data sets**
  - Make data uniform and Merge
  - Scatterplot Independent variables vs. Incidence of Lyme Disease

- **Countywide data sets**
  - Make data uniform and Merge
  - Scatterplot Independent variables vs. Lyme Disease Cases

- **Animal Data Sets**
  - Clean data
  - Plot locations of each animal on US Map
Exploration of Plots and Best Predictors

- Given the initial data our client provided, we were able to see which areas had the highest number of tick cases
- Majority of cases were found in Northeast portion of the United States
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5
Choropleth Graphs and Final Findings

- Bivariate Choropleth Graph of Temperature and Precipitation
- Bivariate Choropleth Graph of Total Population and Population Density
Precipitation and Temperature vs. Lyme Disease Cases

U.S. States' Average Temperature and Total Precipitation
This is a bivariate choropleth map of the United States

1 dot placed randomly within county of residence for each confirmed case
Population Density and Total vs. Lyme Disease Cases
What We’ve Learned/Future Work

● What We’ve learned
  ○ Timeline and scheduling
  ○ Problems faced
    ■ Had to settle for state
    ■ R had to be learned

● Future Work
  ○ Updating data
  ○ Make a website
Acknowledgements

Client: Dr. Escobar

Other Acknowledgements:

- Dr. Fox
- Mariana Guzman
References

Data Collected from:

- [https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income](https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income)
- [https://randstatestats.org/](https://randstatestats.org/)

Additional References:

- [http://my.ilstu.edu/~jrcarter/Geo204/Choro/](http://my.ilstu.edu/~jrcarter/Geo204/Choro/)
- [https://www.cdc.gov/lyme/index.html](https://www.cdc.gov/lyme/index.html)