Our contingency table...
```
                Deaths          Non Deaths
Close to sewer  224             482
Far from sewer  436             1047
```

```python
# Now let us get our p-value!
# ... when doing data science in Python, it is common convention to use
#.    "_" characters to mark variables whose values we don't need.

from scipy.stats import chi2_contingency
chi_square, p_value, _, _ = chi2_contingency(contingency_table)
print(f"Our p-value: {p_value}")
print(f"Our Chi-squared value: {chi_square}")
```

```
Our p-value: 0.28927858447134214
Our Chi-squared value: 1.1229758183695626
```

Death rate per 1000
North: 1.7805598480785025
South: 12.59968677572774
East: 6.326962726552367
West: 3.2989610482397667
Central: 5.021616246078257

Year	Population	Deaths	Deaths per 1000
1840	1842458	46281	25.119161
1841	1877963	45284	24.113361
1842	1916860	45272	23.617792
1843	1953787	48574	24.861461
1844	2033816	50423	24.792312
1845	2073298	48332	23.311651
1846	2113535	49089	23.226017
1847	2195401	60442	27.531189
1848	2238703	57628	25.741691
1849	2282858	68432	29.976459
1850	2327884	48579	20.868308
1851	2373799	55354	23.318739
1852	2420619	54213	22.396337
1853	2468362	61202	24.794580
1854	2517048	73697	29.279140

Our contingency table...
```
             Deaths 1849      Deaths 1854
Mostly SV    3708             3726.0
Mostly Lam   2108             895.0
```

Our contingency table...
```
               Deaths          Non Deaths
Close to pump  492             214
Far from pump  303             1180
```

```python
# Task 3: Determine if there is a statistically signif

from scipy.stats import chi2_contingency

# Now let us get our p-value!
# ... when doing data science in Python, it is common
#.    "_" characters to mark variables whose values we
chi_square, p_value, _, _ = chi2_contingency(contingen
print(f"Our p-value: {p_value}")
print(f"Our Chi-squared value: {chi_square}")
```

```
Our p-value: 1.2325783049887838e-79
Our Chi-squared value: 357.0551926130695
```
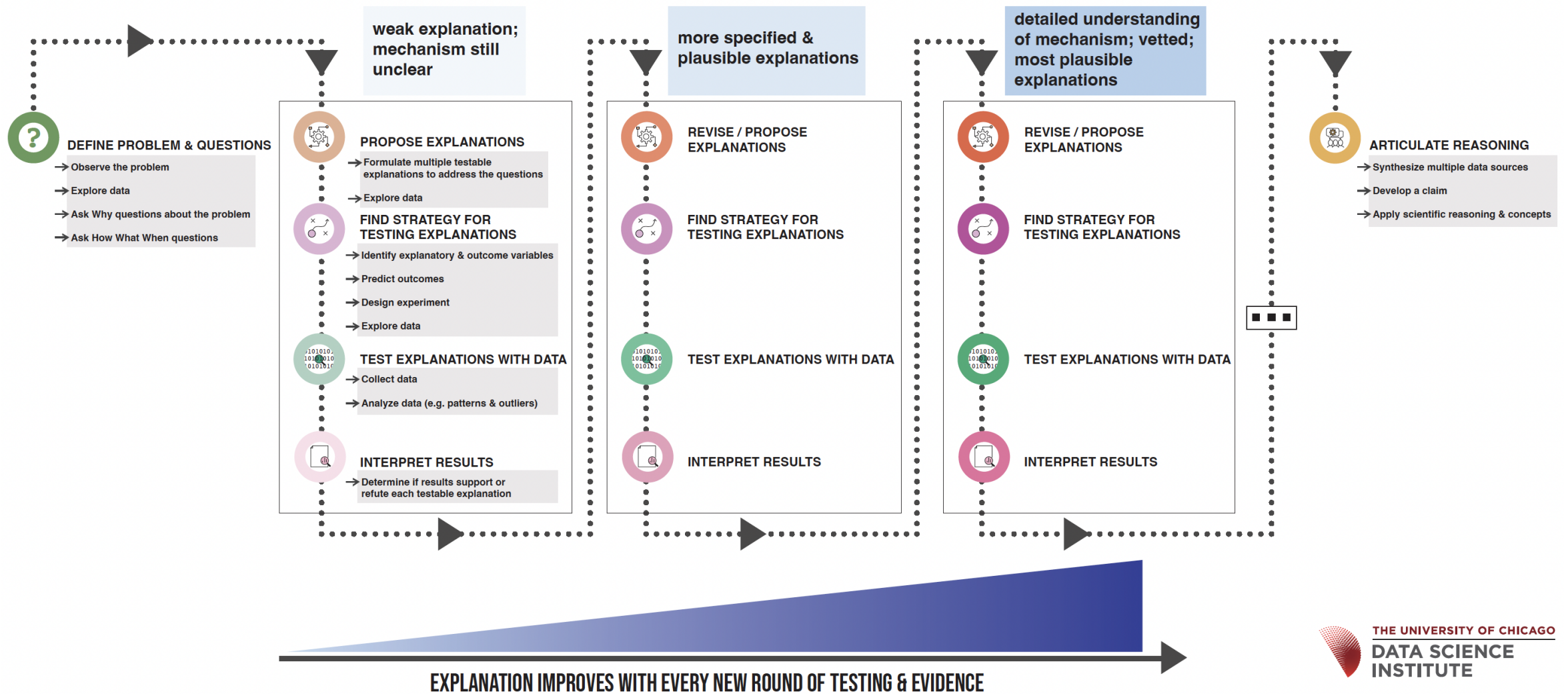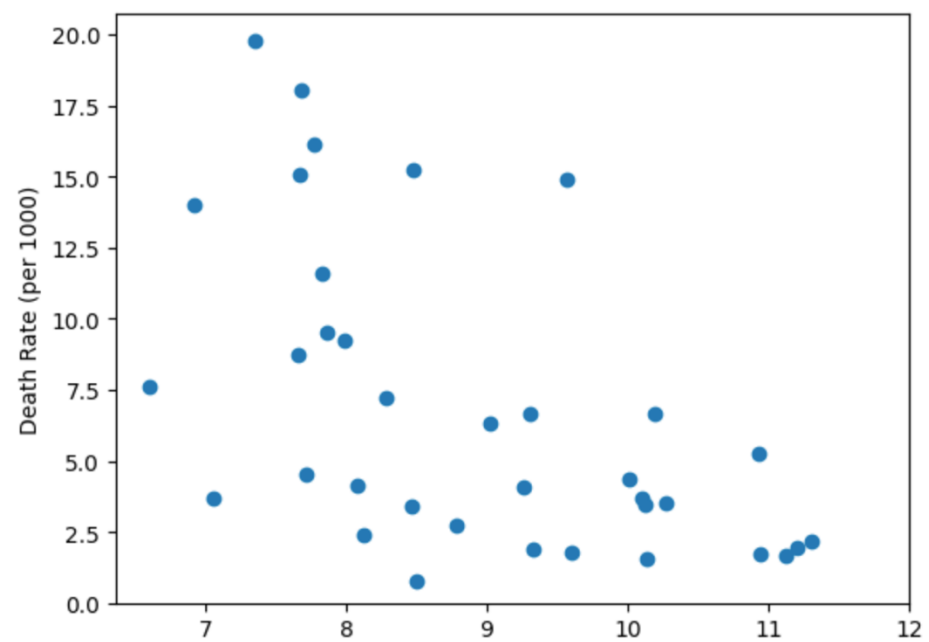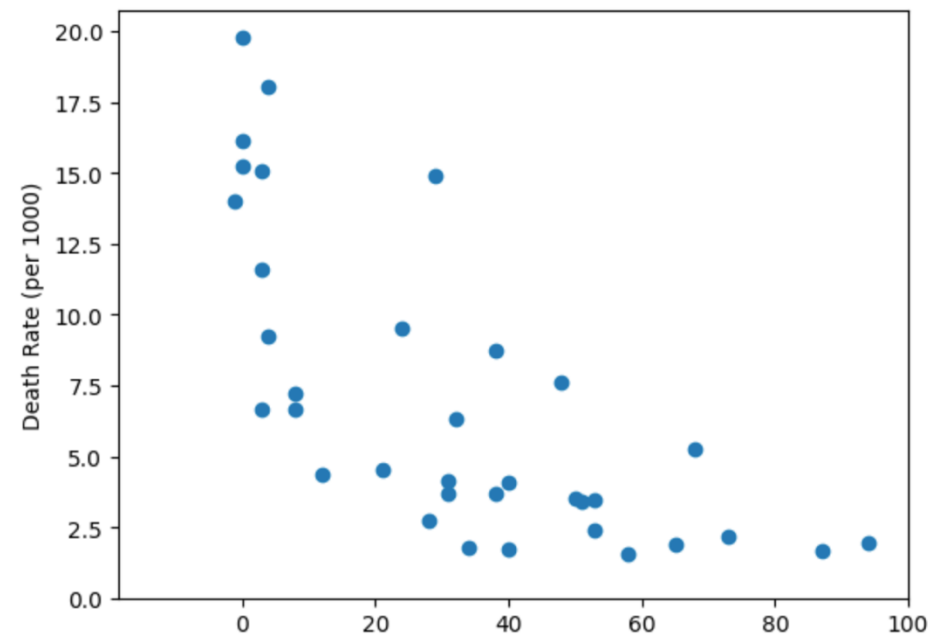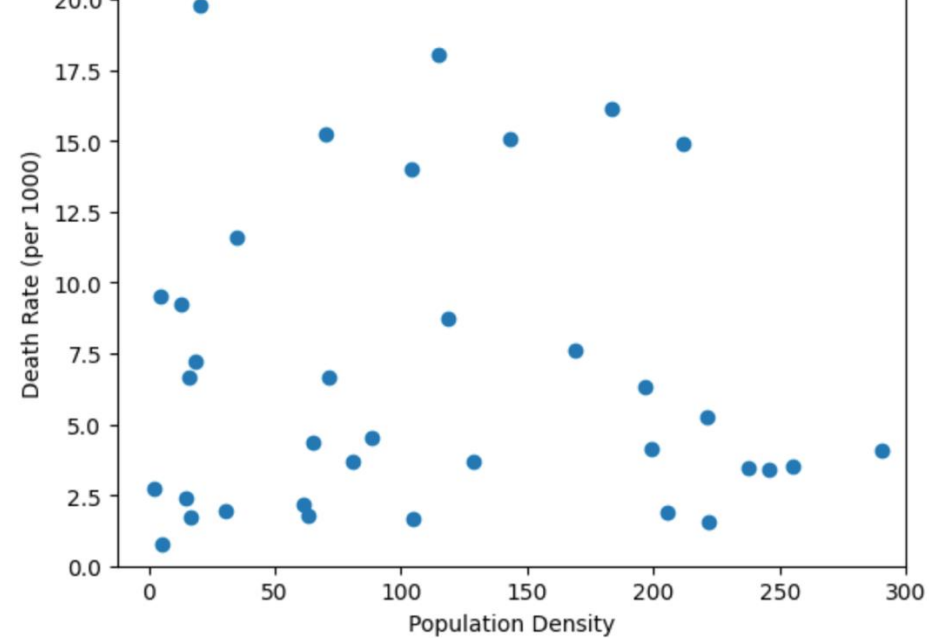
```python
# Find p-value here.

# Now let us get our p-value!
# ... when doing data science in Python, it is common convention to use
#.    "_" characters to mark variables whose values we don't need.

from scipy.stats import chi2_contingency
chi_square, p_value, _, _ = chi2_contingency(contingency_table)
print(f"Our p-value: {p_value}")
print(f"Our Chi-squared value: {chi_square}")
```

```
Our p-value: 1.1436488037757276e-110
Our Chi-squared value: 499.6308374196881
```

THE DATA SCIENCE REASONING FRAMEWORK

DEFINE PROBLEM & QUESTIONS
→ Observe the problem
→ Explore data
→ Ask Why questions about the problem
→ Ask How What When questions

weak explanation; mechanism still unclear

PROPOSE EXPLANATIONS
→ Formulate multiple testable explanations to address the questions
→ Explore data

FIND STRATEGY FOR TESTING EXPLANATIONS
→ Identify explanatory & outcome variables
→ Predict outcomes
→ Design experiment
→ Explore data

TEST EXPLANATIONS WITH DATA
→ Collect data
→ Analyze data (e.g. patterns & outliers)

INTERPRET RESULTS
→ Determine if results support or refute each testable explanation

more specified & plausible explanations

REVISE / PROPOSE EXPLANATIONS

FIND STRATEGY FOR TESTING EXPLANATIONS

TEST EXPLANATIONS WITH DATA

INTERPRET RESULTS

detailed understanding of mechanism; vetted; most plausible explanations

REVISE / PROPOSE EXPLANATIONS

FIND STRATEGY FOR TESTING EXPLANATIONS

TEST EXPLANATIONS WITH DATA

INTERPRET RESULTS

ARTICULATE REASONING
→ Synthesize multiple data sources
→ Develop a claim
→ Apply scientific reasoning & concepts

EXPLANATION IMPROVES WITH EVERY NEW ROUND OF TESTING & EVIDENCE

THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE

A COURT FOR KING CHOLERA.

Deaths and Nondeaths (Close and Far from Sewers): HISTOGRAM

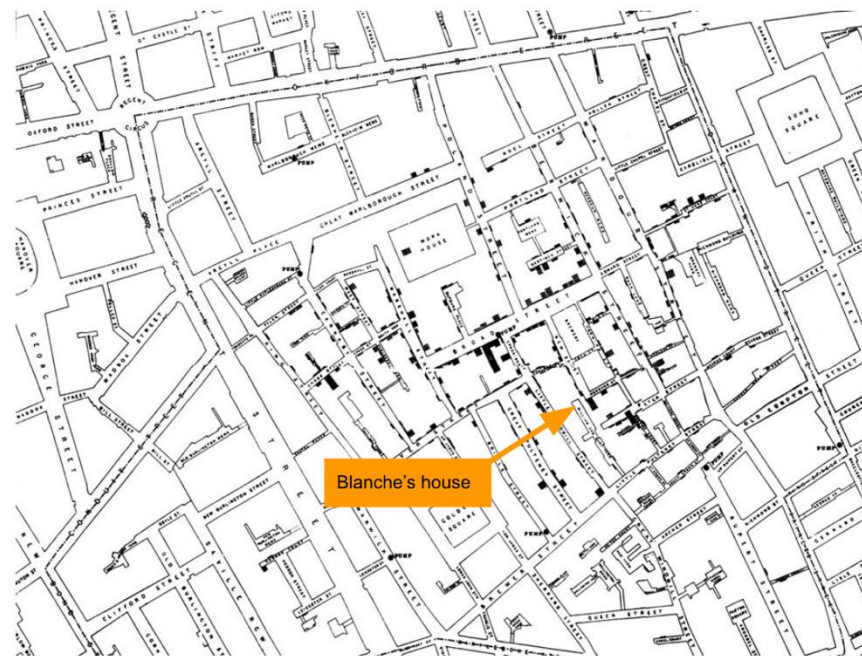**Deaths and Nondeaths (Close and Far from Broad St. Pump): HISTOGRAM**
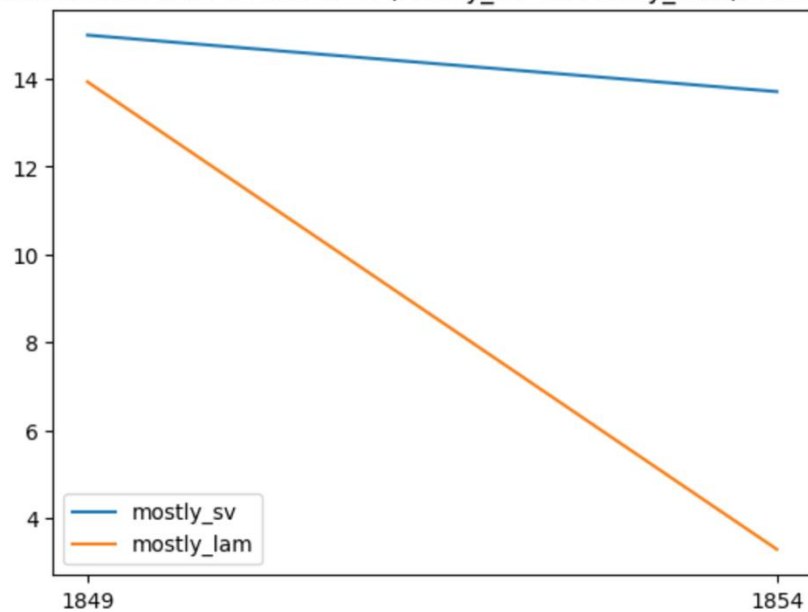
Cholera deaths of 1849 and 1854 (Mostly_SV vs. Mostly_Lam): LINE GRAPH
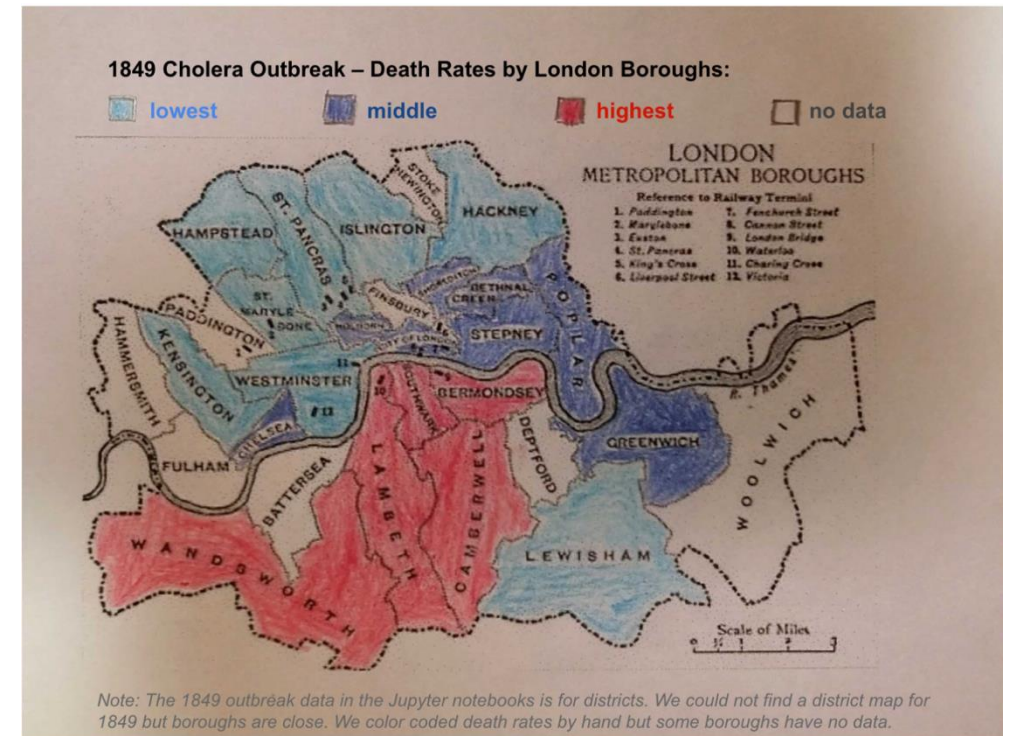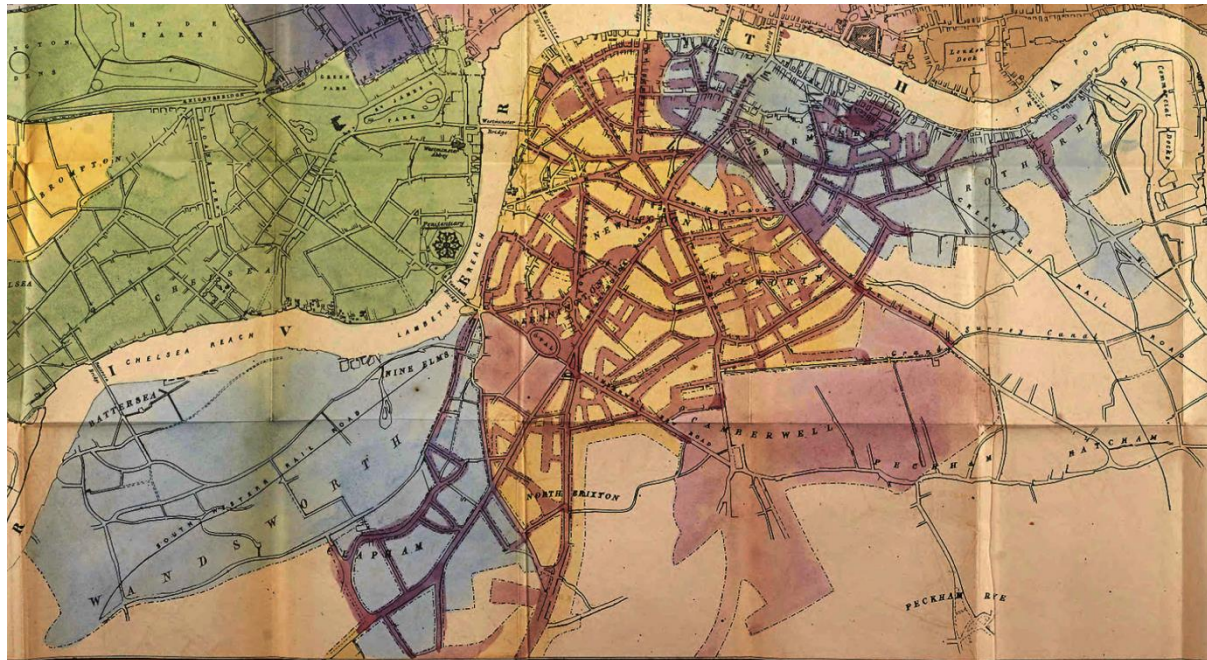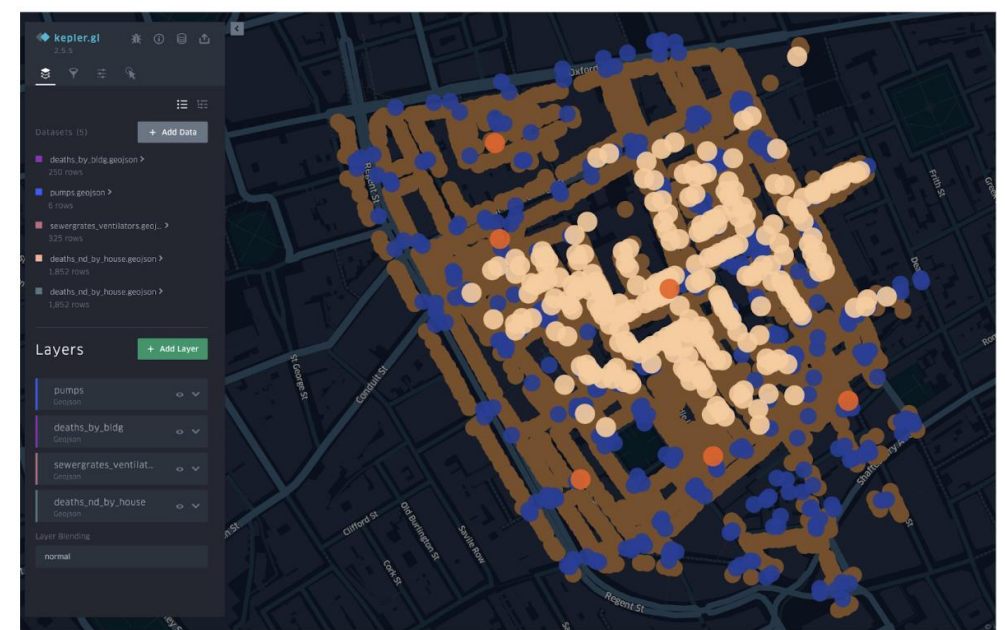
Blanche's house

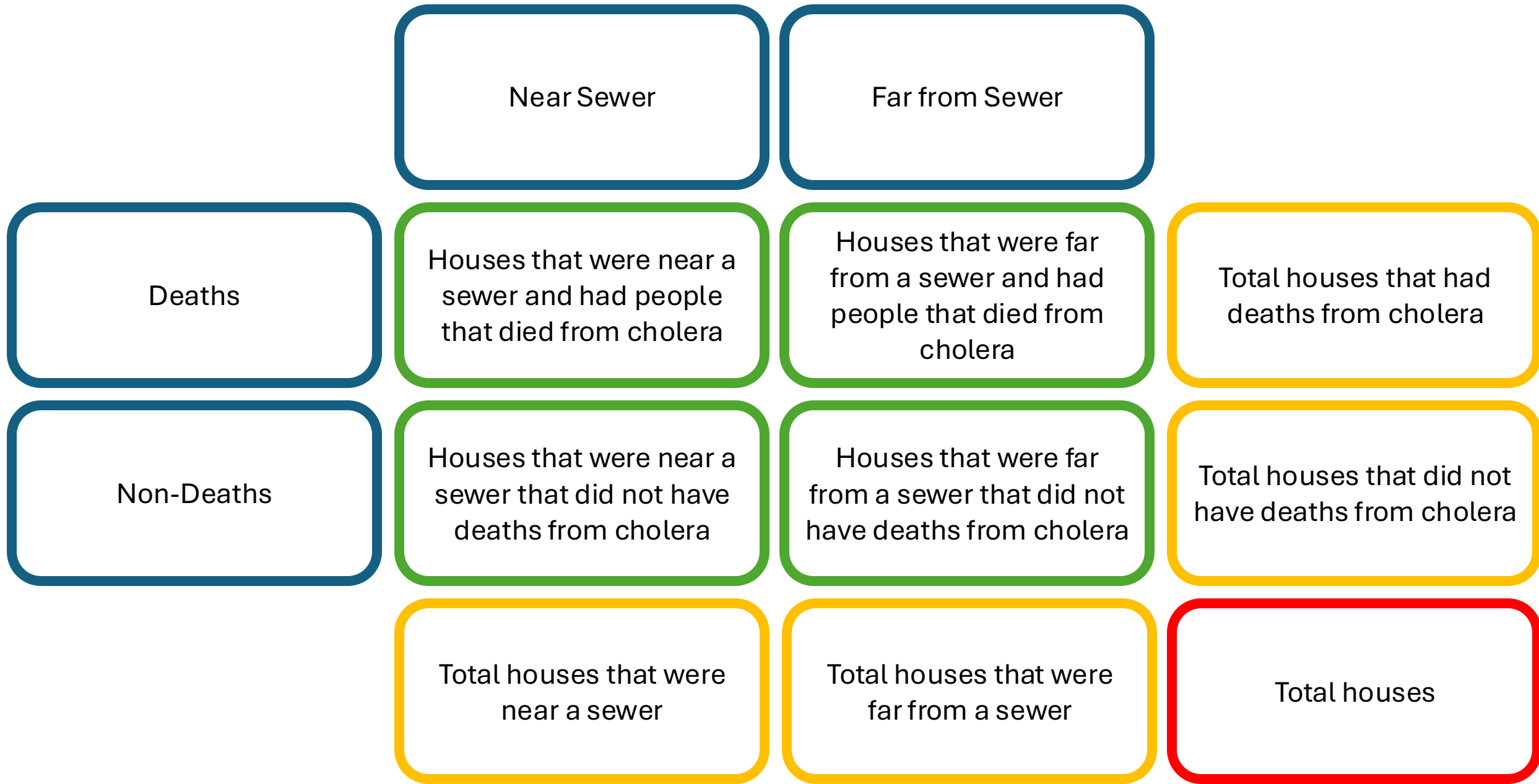At 7, Layton's Buildings, July 29, a tailor, aged 20, "cholera 17 hours"   .   *Southwark and Vauxhall.*

At 2, Dobb's Cross, July 30, the son of a shop-keeper, aged 10 yrs., "cholera Asiatic 24 hours". *Southwark & Vauxhall.*

At 81, Ann Street, July 29, the son of a labourer, aged 12 years, "cholera 8 hours".   *Southwark and Vauxhall.*

At 28, Wickham Pl., Aug. 2, son of a brush-mkr., aged $2\frac{1}{2}$ yrs., "choleraic diarrhœa 24 hours." *Southwark & Vauxhall.*

At 2, Russell Place, Aug. 2, the widow of a labourer,

1849 Cholera Outbreak – Death Rates by London Boroughs:

lowest   middle   highest   no data

Note: The 1849 outbreak data in the Jupyter notebooks is for districts. We could not find a district map for 1849 but boroughs are close. We color coded death rates by hand but some boroughs have no data.

	Near Sewer	Far from Sewer	
Deaths	Houses that were near a sewer and had people that died from cholera	Houses that were far from a sewer and had people that died from cholera	Total houses that had deaths from cholera
Non-Deaths	Houses that were near a sewer that did not have deaths from cholera	Houses that were far from a sewer that did not have deaths from cholera	Total houses that did not have deaths from cholera
	Total houses that were near a sewer	Total houses that were far from a sewer	Total houses

	Near the Broad Street Pump (BSP)	Far from the Broad Street Pump (BSP)	
Deaths	Houses that were near the BSP and had people that died from cholera	Houses that were far from the BSP and had people that died from cholera	Total houses that had deaths from cholera
Non-Deaths	Houses that were near the BSP that did not have deaths from cholera	Houses that were far from the BSP that did not have deaths from cholera	Total houses that did not have deaths from cholera
	Total houses that were near the BSP	Total houses that were far from the BSP	Total houses

	Subdistricts supplied mostly by Lambeth	Subdistricts supplied mostly by Southwark & Vauxhall
Deaths in 1849	Deaths in subdistricts supplied by Lambeth in 1849	Deaths in subdistricts supplied by Southwark & Vauxhall in 1849
Deaths in 1854	Deaths in subdistricts supplied by Lambeth in 1854	Deaths in subdistricts supplied by Southwark & Vauxhall in 1854