Chlamydia detection rate for young adults aged 15-24 in England (2012 – 2020)

**Introduction**

Chlamydia is a common STI that can infect both men and women. Untreated chlamydia can cause serious complications such as reproductive ill-health, symptomatic acute infections and complications such as pelvic inflammatory disease (PID), ectopic pregnancy and tubal-factor infertility. All of this is avoidable if regular screenings are done.

Rates in England are substantially higher in young adults than any other age group, but could this be due to socio-economic disparities evident across England? In this dataset I will explore the relation between chlamydia detection and area depravity.

**What is Index of Multiple Deprivation?**

The Index of Multiple Deprivation (IMD) combines information from 7 domains to produce an overall relative measure of deprivation. The domains are:

* Income
* Employment
* Education
* Skills and Training
* Health and Disability
* Crime
* Barriers to Housing Service
* Living Environment

IMD ranks every small area in England from 1 (most deprived area) to 32,844 (least deprived area). The IMD’s in this dataset were recorded in 2015 and 2019.

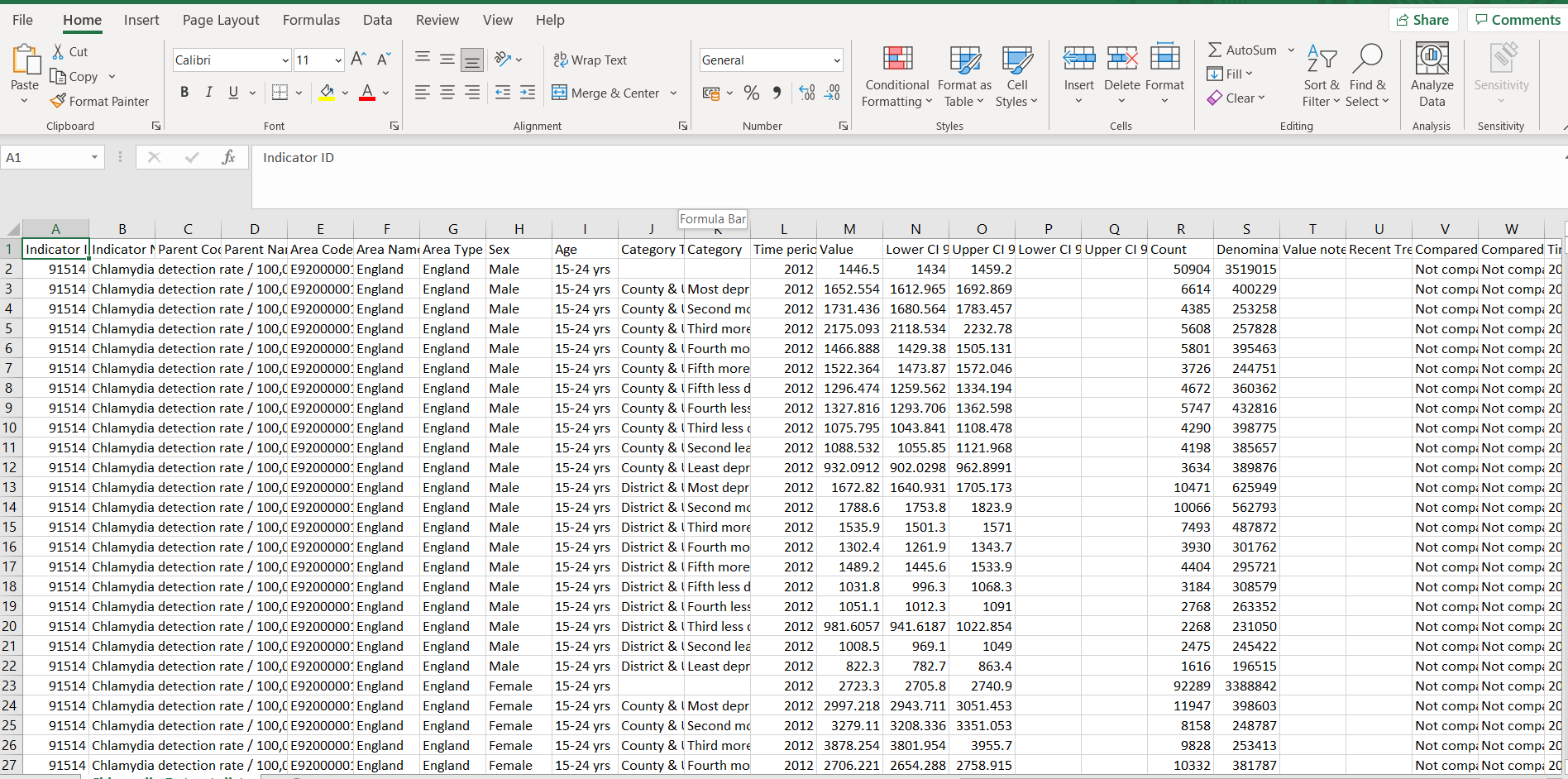
**Dataset**

This dataset is provided by public health England. The data shows chlamydia diagnoses in 15 to 24 year olds attending sexual health services (SHSs) and community-based settings, who are residents in England, expressed as a rate per 100,000 population. Data exclude people accessing services located in England who are residents in Wales, Scotland, Northern Ireland or abroad**.**

**Data cleaning**

The initial data wasn’t super messy but still messy nevertheless. I started first with a pre-analysis of the dataset, which seemed complete (mostly) consistent and accurate. I then proceeded to remove whole columns with null or blank data, looked for any typos or misspellings and removed duplicates using excels duplicate removal tool.

I finally removed any irrelevant columns as well as removing the year of the IMD as it is not relevant for the purposes of the problem question. After doing this I realised the qualitative data in the IMD column would be troublesome to sort and query later down the line, so I decided to attribute them to numbers using the excel find and replace tool. The key for these numbers can be found in the 2nd image below.



Table

Description automatically generated

Table

Description automatically generated with medium confidenceI removed the “yrs” from the age column as its unnecessary characters.

**Data Analysis/exploration**

Firstly, because I will be looking at the IMD value I will need to aggregate the counts from areas with the same IMD in the same year – for this I will have to separate male and female as this may be a variable that has an effect on the correlation between area depravity and Chlamydia detection. I did this in excel and uploaded the two new tables into postgreSQL as Chlamydia\_F and Chlamydia\_M along with the complete data.

**Null hypothesis:** *“There is no relationship between the index of multiple deprivation and the detection of Chlamydia in 15-24 year olds”*

Using the query tool, I started to look at each year by aggregating the counts (Total\_count) for the same IMDs in the year. At a glance we can see that the most deprived areas have the highest counts for chlamydia detection in both males and females, though females have an average of 53% higher detection rates in 2012. This is the general trend until 2020 where the 3rd more deprived decile has the highest detection.

--Query used to aggregate the counts of the same IMD in 2012 for males

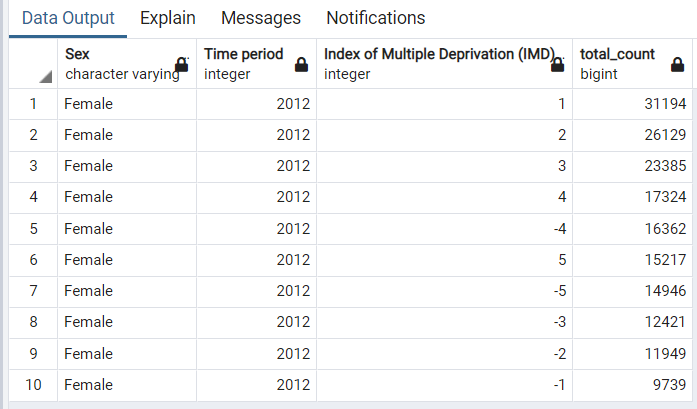
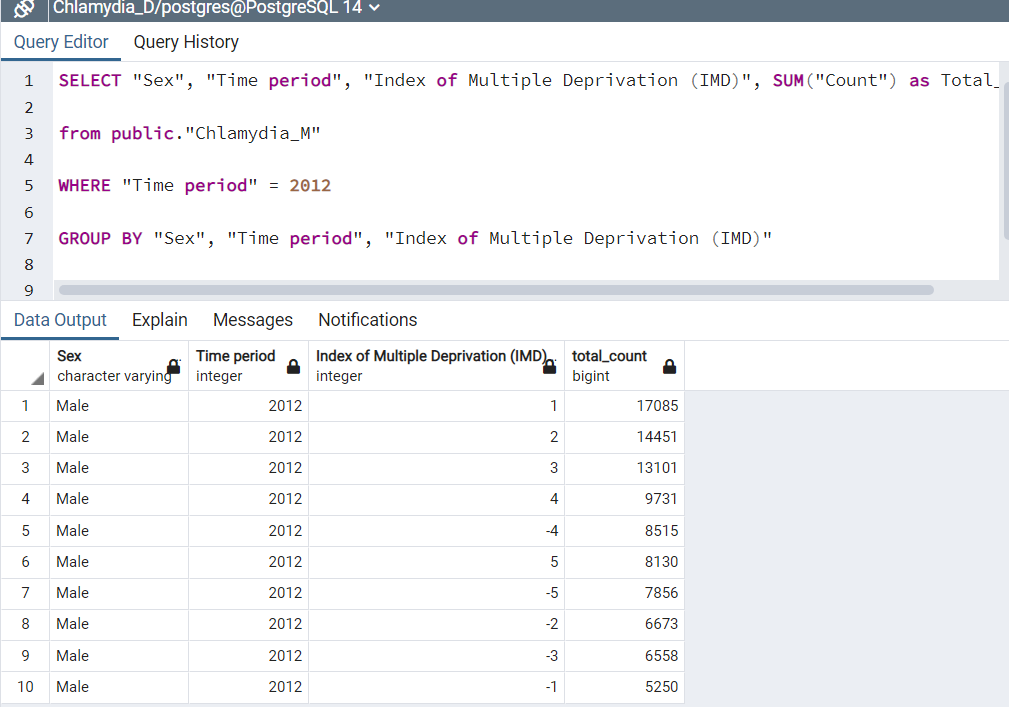
SELECT "Sex", "Time period", "Index of Multiple Deprivation (IMD)", SUM("Count") as Total\_count

from public.Chlamydia\_M

WHERE "Time period" = 2012

GROUP BY "Sex", "Time period", "Index of Multiple Deprivation (IMD)"

ORDER BY total\_count DESC



--Total count of chlamydia detection in every year for each IMD for males

SELECT "Sex", "Time period", "Index of Multiple Deprivation (IMD)", SUM("Count") as Total\_count

from public.Chlamydia\_M

GROUP BY "Sex", "Time period", "Index of Multiple Deprivation (IMD)"

ORDER BY "Index of Multiple Deprivation (IMD)", "Time period" DESC

To make it easier for a “client” or anyone to retrieve data from each year, I have created view tables that have both female and male total counts - in descending order - of chlamydia detection.

CREATE VIEW chlamydia\_2012 AS

SELECT "Sex", "Time period", "Index of Multiple Deprivation (IMD)", SUM("Count") as Total\_count

from public.Chlamydia

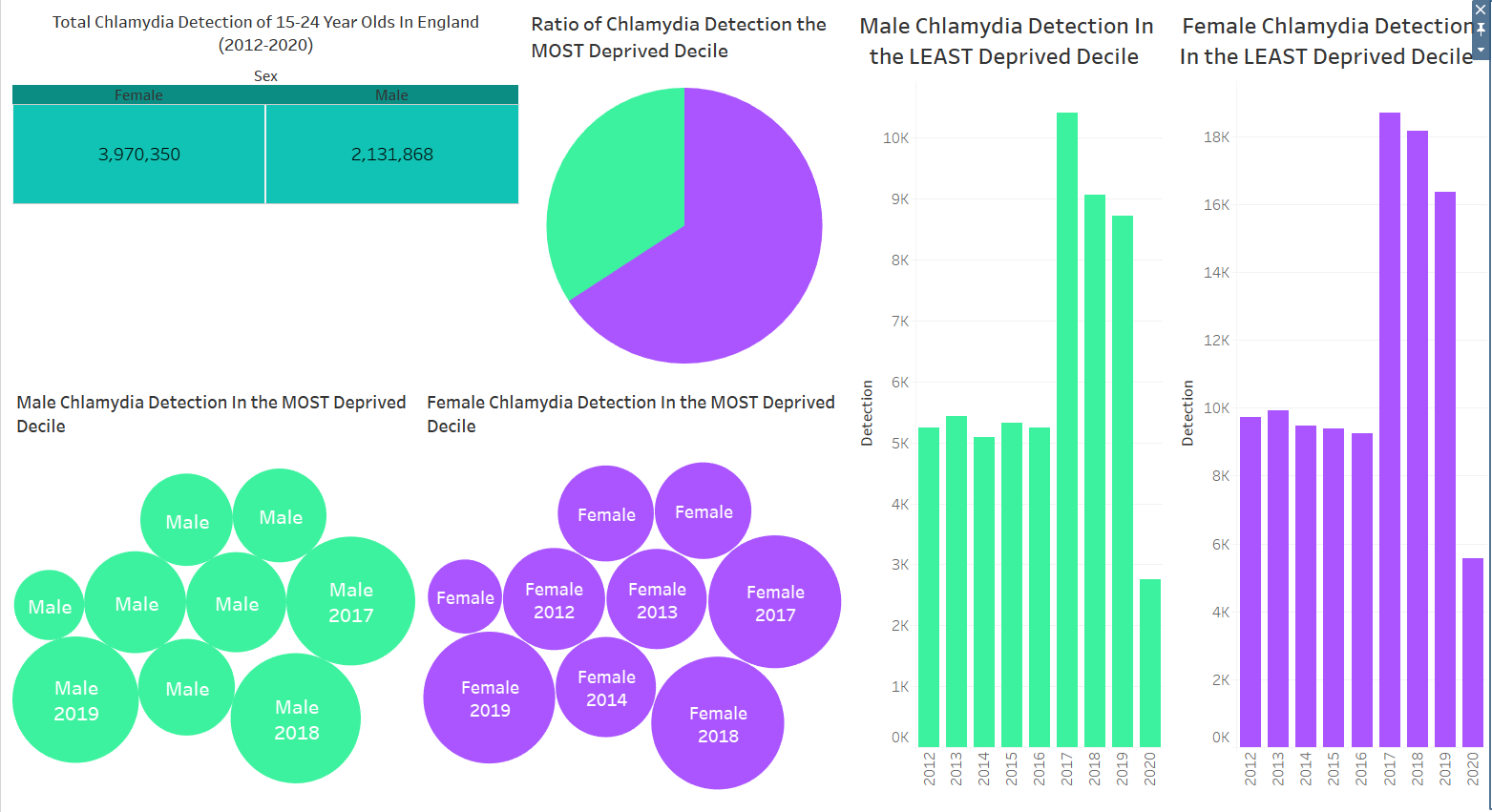
WHERE "Time period" = 2012 – every year has it own view table

GROUP BY "Sex", "Time period", "Index of Multiple Deprivation (IMD)"

ORDER BY "Sex", total\_count desc

**Data visualisation using Tableau**

Because the data rejects our null hypothesis, I see it suitable to focus on the most deprived decile (1) and the least deprived decile (-1) - though i will show an overview on my dashboard to tell a more well-rounded story.



I used these specific colours for male and female as they are gender fluid.

**Link to data visualisation**: <https://public.tableau.com/app/profile/tiyana.kelly>

**Theories**

*Why the detection levels high in 2017-2019*

After research, I have found that in 2017 there were notable changes in sexual health service provision in England – mainly the introduction of internet services and tech. Gen Z and millennials are a generation of convenience and tech/social media dependency so making STI testing more accessible would increase the testing for these age groups

*Why were the detection levels so low in 2020?*

2020 was the year that COVID had spread to the UK. Beginning of march, lockdown rules were put in place across England and non-essential hospital visits were limited and, in some cases, not allowed. Sexual health clinics would have been running at a very limited compacity. Which could explain the low counts.

Also, you have to take into account the behavioural aspect of a national lockdown. As a result, there would have been far less overall physical human interactions - especially in prime youth ‘mating’ grounds such as clubs and bars as well as tinder hook ups.

*Why is there almost double the detection count for females compared to males*

I don’t want to be *that* person but women are more likely to go to the sexual health clinic or GP than men [1], so this could the source of why the detection rate is higher in women compared to men.

**Citations & References**

[1] The Conversation Health Journal - https://theconversation.com/why-women-see-their-gp-more-than-men-49051#:~:text=Studies%20show%20continuity%20of%20care,had%20a%20regular%20GP%20practice.