

# AM 170B HW 1

Henry Charlton

2022 April 6

## 1 COVID-19 United States Analysis

### 1.1 Data

Data for this analysis was retrieved from the U.S. CDC website as a csv file, with the format of several variables pertaining to the COVID-19 pandemic in a time-series basis for each state. The variables used in this report were new cases, new deaths, state, total deaths, and submission date. Data manipulation was done primarily using the pandas package in python 3.

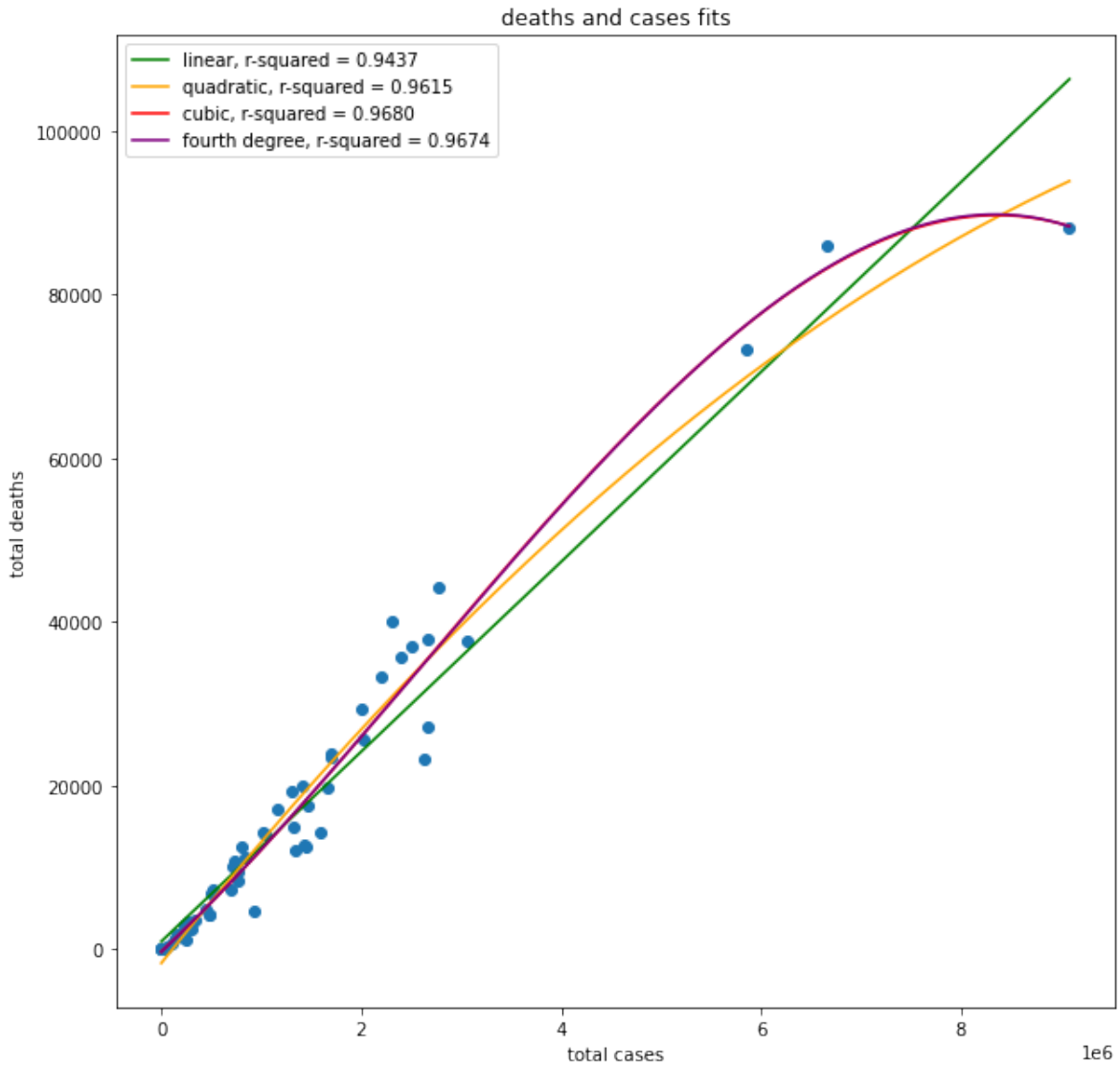


Figure 1: Deaths by Cases and Various Fits

## 1.2 Analysis

In figure 1 we notice that the overall shape of the mapping between cases and deaths is mostly linear. The r-squared error for the linear model is only slightly worse than for the polynomials, and since it has the added benefit of simplicity we conclude that it is the most reasonable fit.

Figure 2 demonstrates the overall distribution of new cases for the most recent day recorded, per state. We note that most of the states had under 2000 cases, with only a few outliers including California.

Figure 3 allows us to compare several chosen states and their time series of new cases per day over the entire temporal span of the dataset.

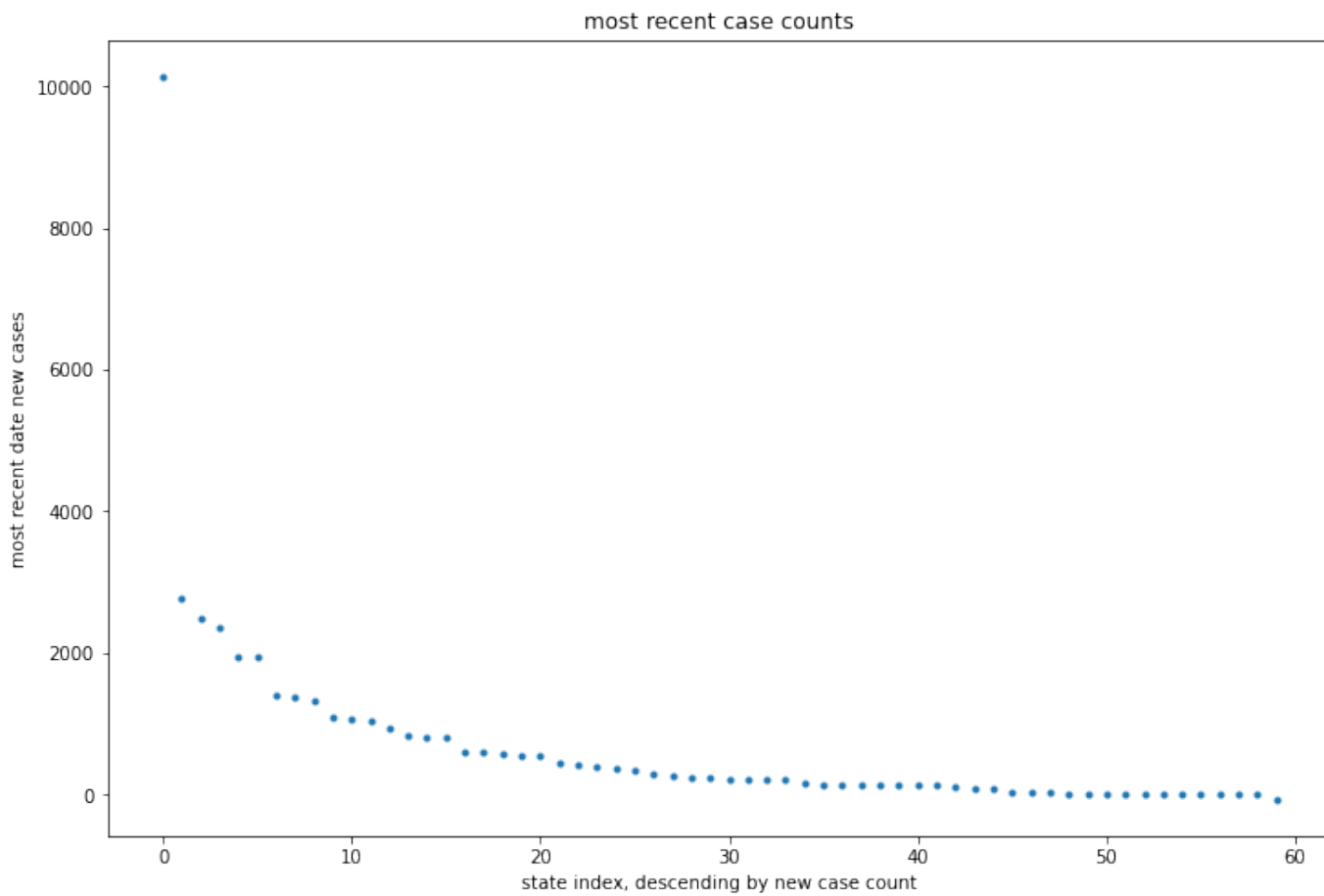


Figure 2: Most Recent Case Count Distribution

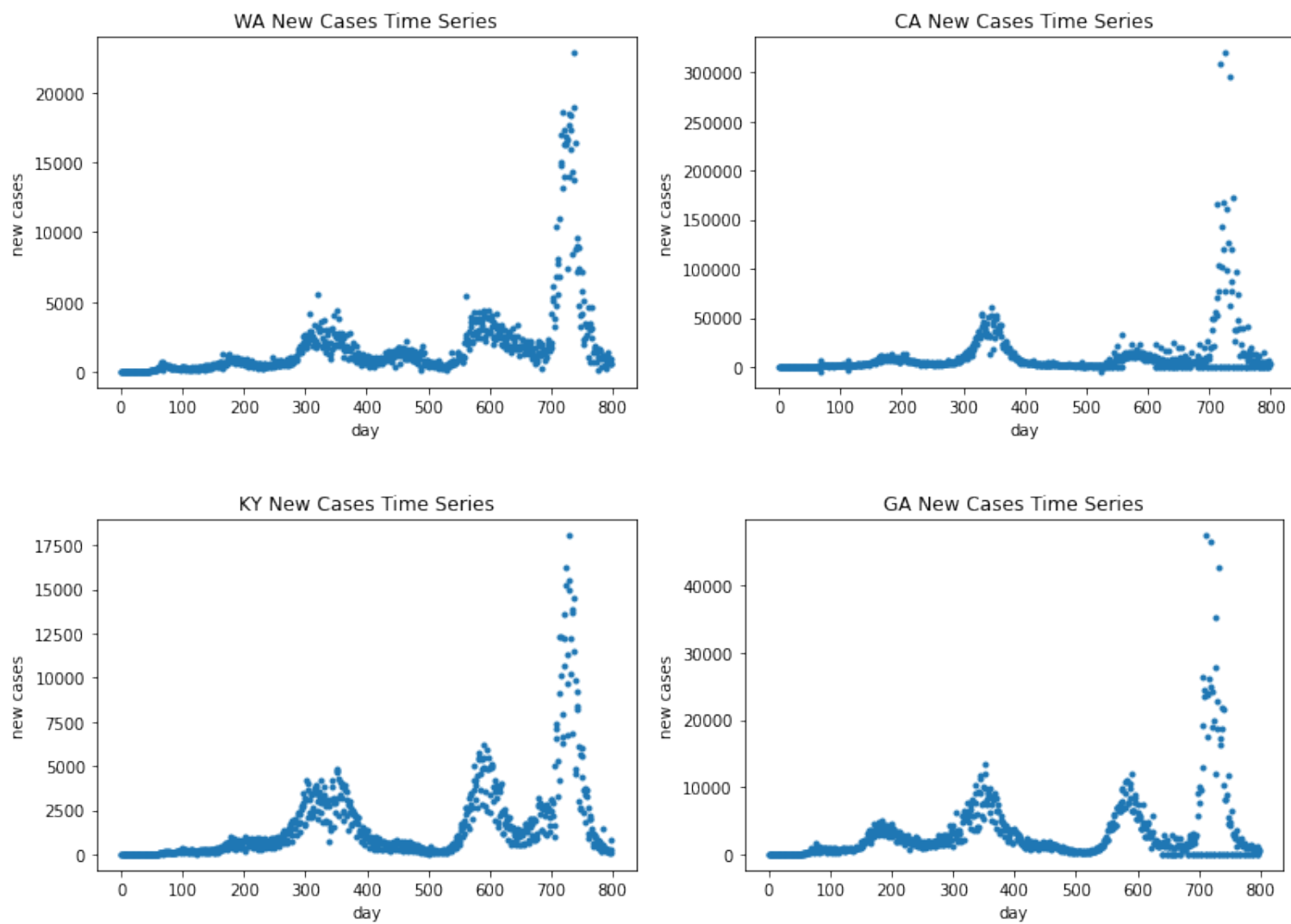


Figure 3: Key States Time Series

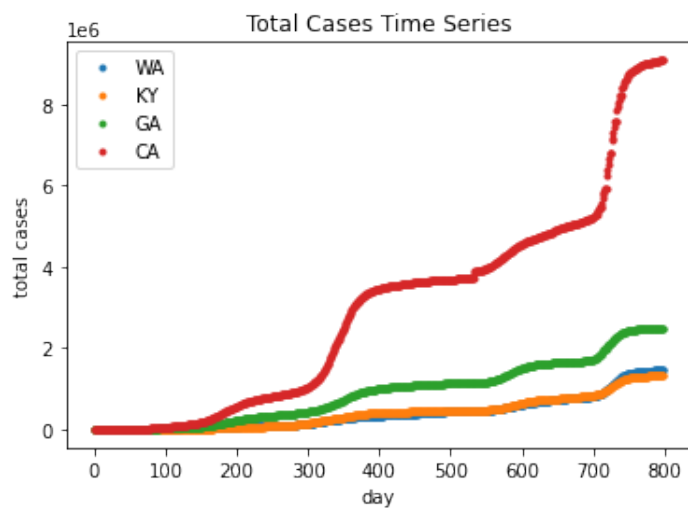


Figure 4: Key States: Total Cases Time Series

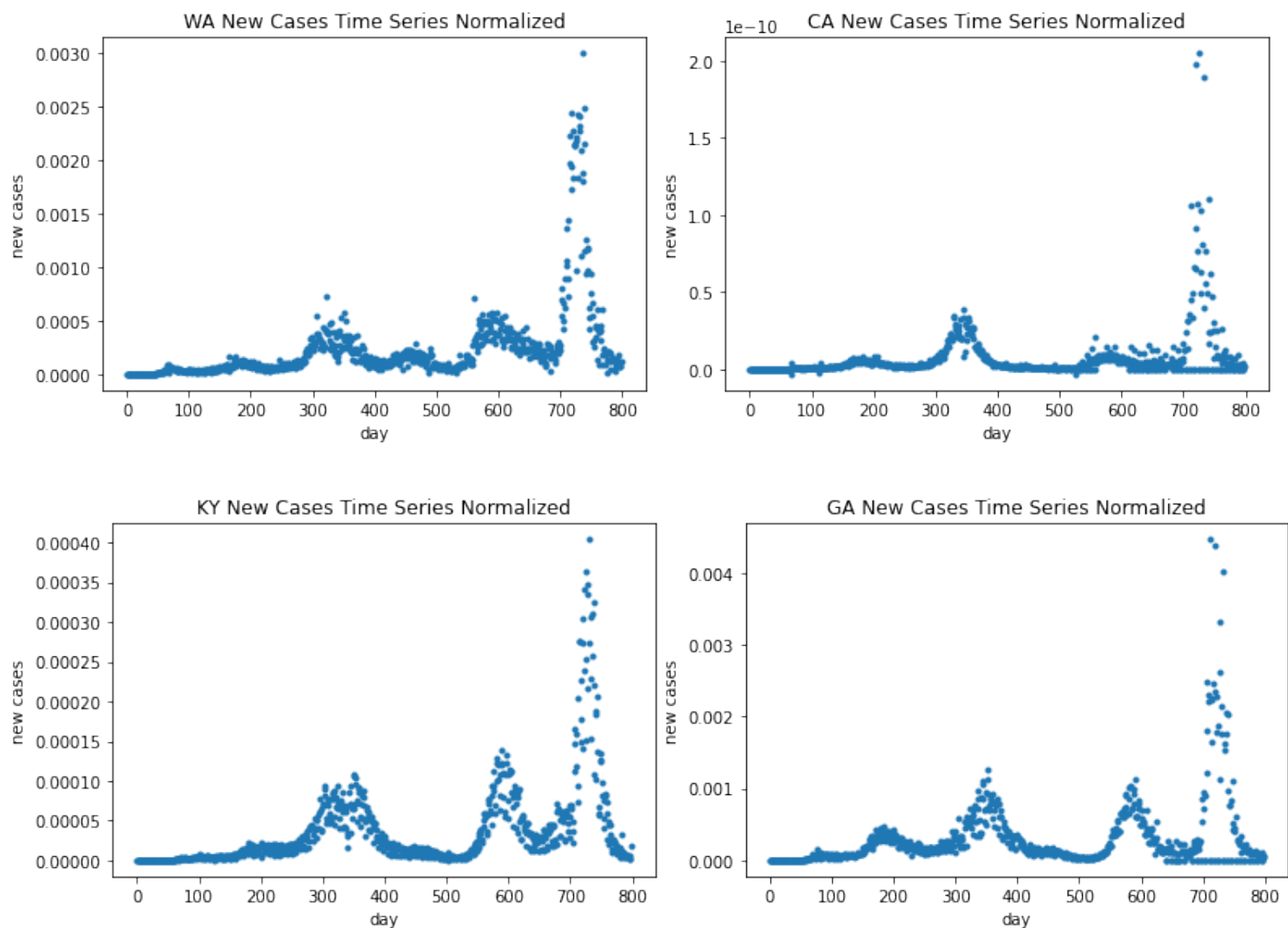


Figure 5: Key States Normalized Time Series

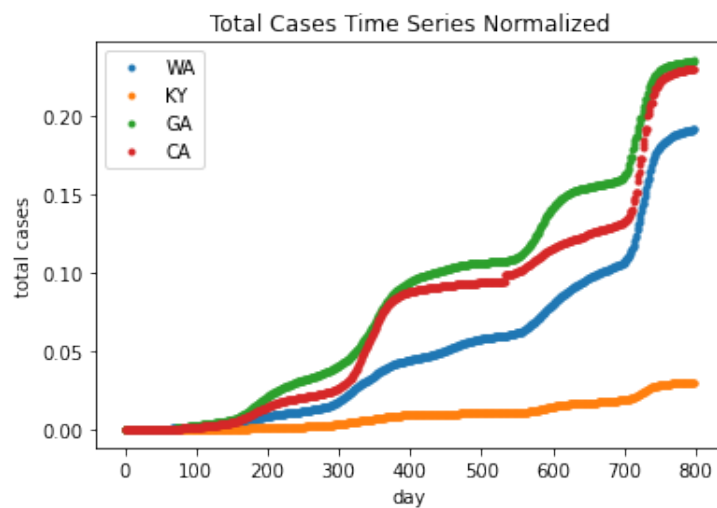
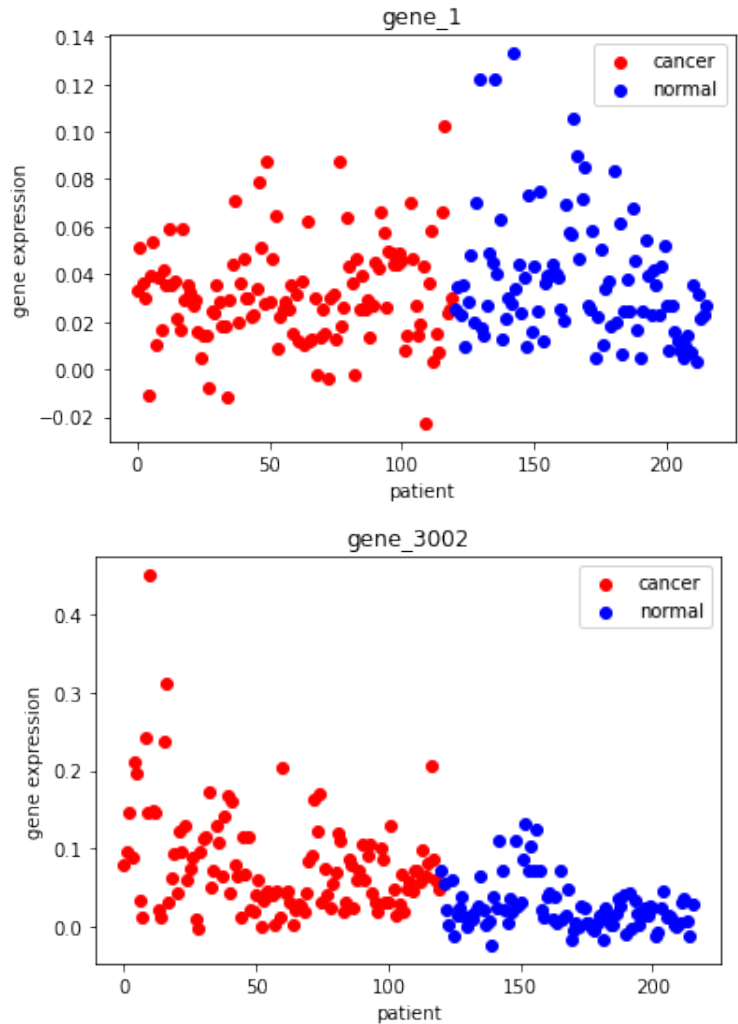


Figure 6: Key States: Total Cases Time Series Normalized

## 2 Ovarian Cancer Analysis



The above figures show the gene expressions for the measured genes, color coded to show which were associated with cancer. Note that 3999 genes were measured. The figure below shows the total counts of cancerous and non-cancerous genes measured. We note that there is decent parity in the counts, which allows us to have a balanced analysis.

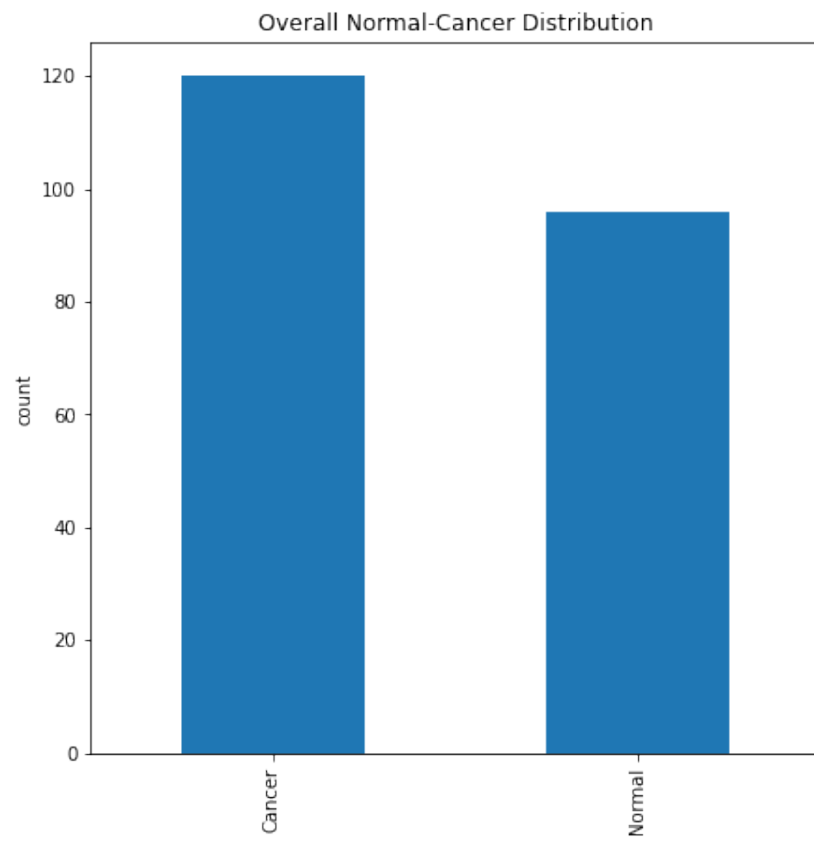


Table 1: log-2 fold change

gene	log-2 fold change
gene 1	-0.238
gene 3002	1.626

Table 2: Mann-Whitney-Wilcoxon test

gene	statistic	p value
gene 1	5373	0.199
gene 3002	2149	$1.282 \cdot 10^{-15}$

Table 3: t-test

gene	statistic	p value
gene 1	1.750	0.082
gene 3002	-7.380	$3.463 \cdot 10^{-12}$

Observing the information in tables 2 and three, we note that by both tests gene 3002 does meet the requirement of a p-value less than 0.05 to reject the null hypothesis, whereas gene 1 does not in either test. It is worth noting however, that the Mann-Whitney-Wilcoxon test got closer to this value, which tells us that it is estimating a higher significance in the difference between the cancer and normal groups of each gene. In both cases however, the cut-off was exceeded by a huge margin, indicating that there is a significant difference between the means of cancer and normal groups for gene 1 and gene 3002.



### 3 Population Dynamics

In the following figures we observe simulations for the predator-prey model for different initial conditions. We note that the dynamics depend on the the initial conditions heavily. Most notably, we observe the in figure 1 the amplitude of the oscillations increases with time whereas in figures 2 and 3 the amplitude of the oscillations settles in consistency. Below is a summary of the system and its equilibrium points:

$$\begin{aligned}\dot{x} &= rx\left(1 - \frac{x}{k}\right) - \frac{axy}{c+x}, x \geq 0 \\ \dot{y} &= \frac{baxy}{c+x} - dy, y \geq 0\end{aligned}$$

Equilibrium points:

$$\begin{aligned}(0, 0) \\ (0, \frac{rc}{a}) \\ (\frac{dc}{ba-d}, \frac{bcr(abk - cd - dk)}{k(ab-d)^2})\end{aligned}$$

Note that in the simulation the following constants are used:

$$a = 3.2, b = 0.6, c = 50, d = 0.56, k = 125, r = 1.6$$

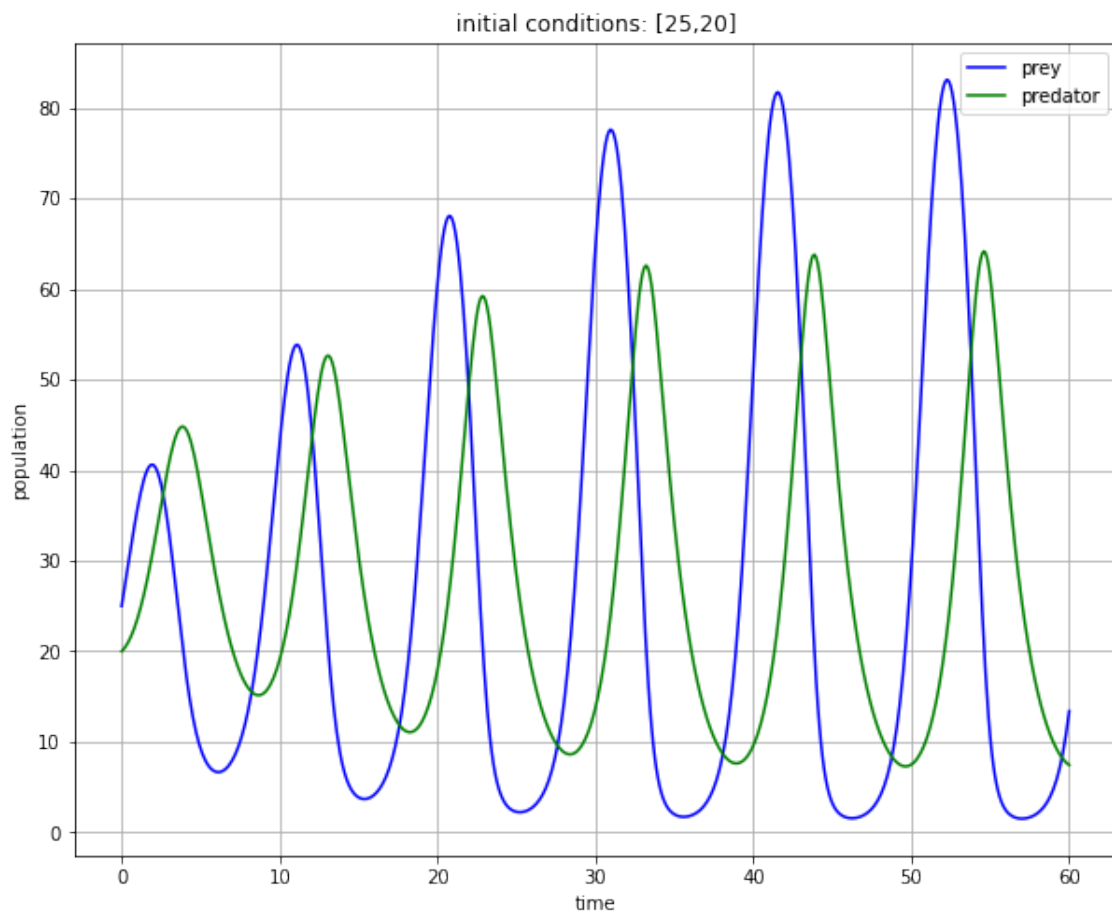


Figure 7: Dynamics: [Prey: 25, Predators: 20]

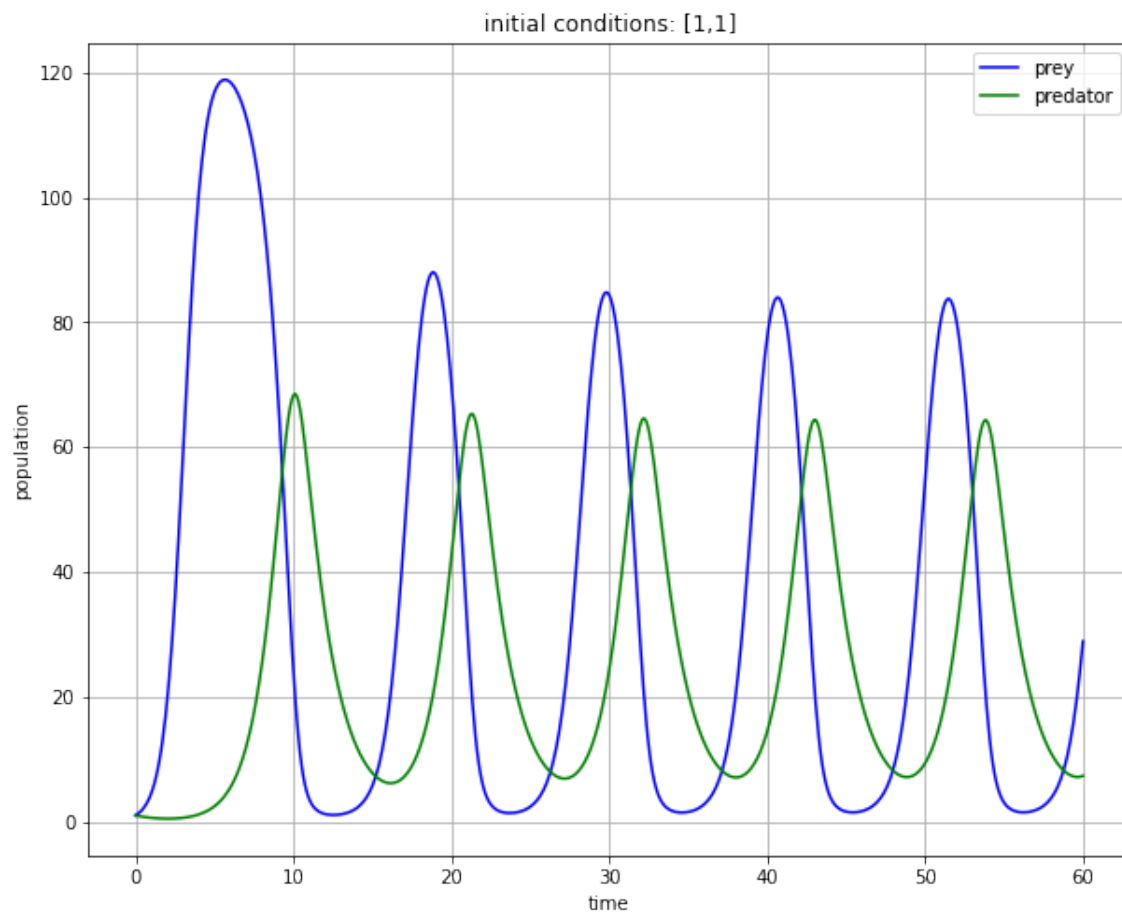


Figure 8: Dynamics: [Prey: 1, Predators: 1]

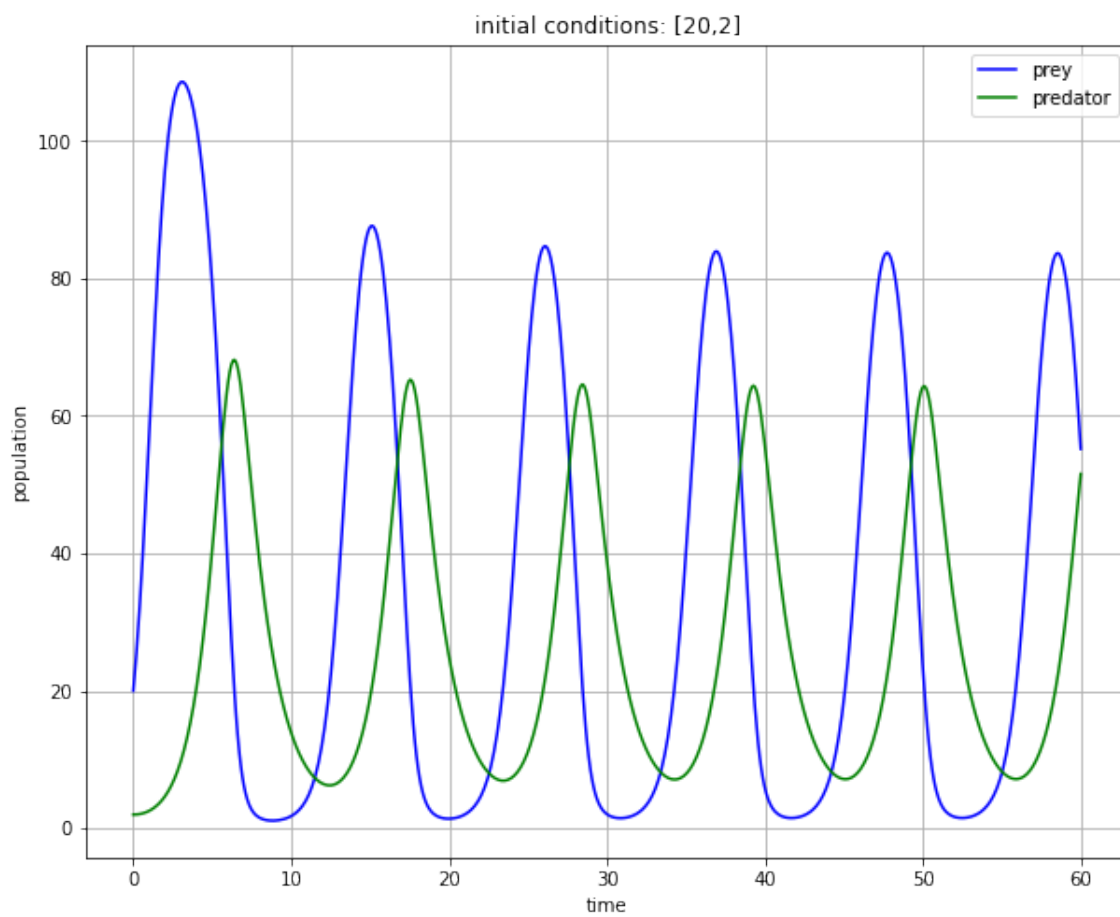


Figure 9: Dynamics: [Prey: 20, Predators: 2]