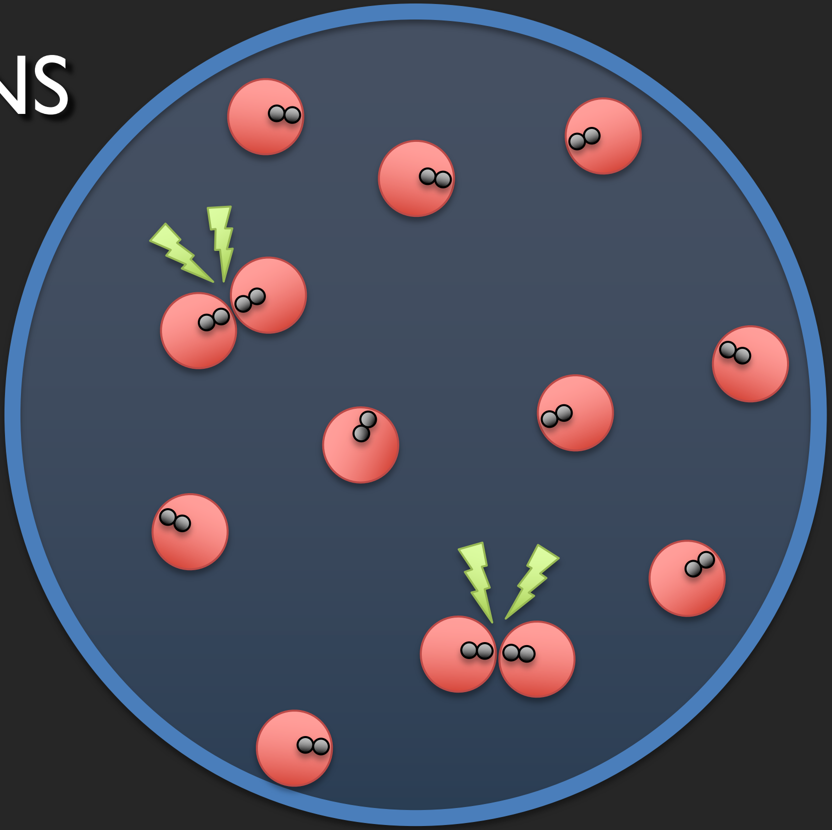
CASA0011: Agent-Based Modelling

Dr Sarah WISE
Dr Thomas OLÉRON EVANS

s.wise@ucl.ac.uk
thomas.evans.11@ucl.ac.uk

Moodle password: ABM2019

Centre for Advanced
Spatial Analysis,
90 Tottenham Court Road



Attendance Sheet

The ABM Course

Week 1: Introduction to ABMs

Week 2: Cellular Automata

Week 3: ABM Methodology

Week 4: Agent Behaviours

Week 5: ABMs as Research Tools

READING WEEK

Week 6: **Testing ABMs / Presenting Results**

Week 7: Modelling Competitive Agents

Week 8: Forecasting & Prediction

Week 9: Traffic Modelling

Week 10:

Course Objectives

You should...

1. understand the principles of agent-based modelling (ABM)

2. be able to describe the type and range of systems to which ABM can be profitably and appropriately applied

3. be able to conceptualise and model urban systems with complex dynamics

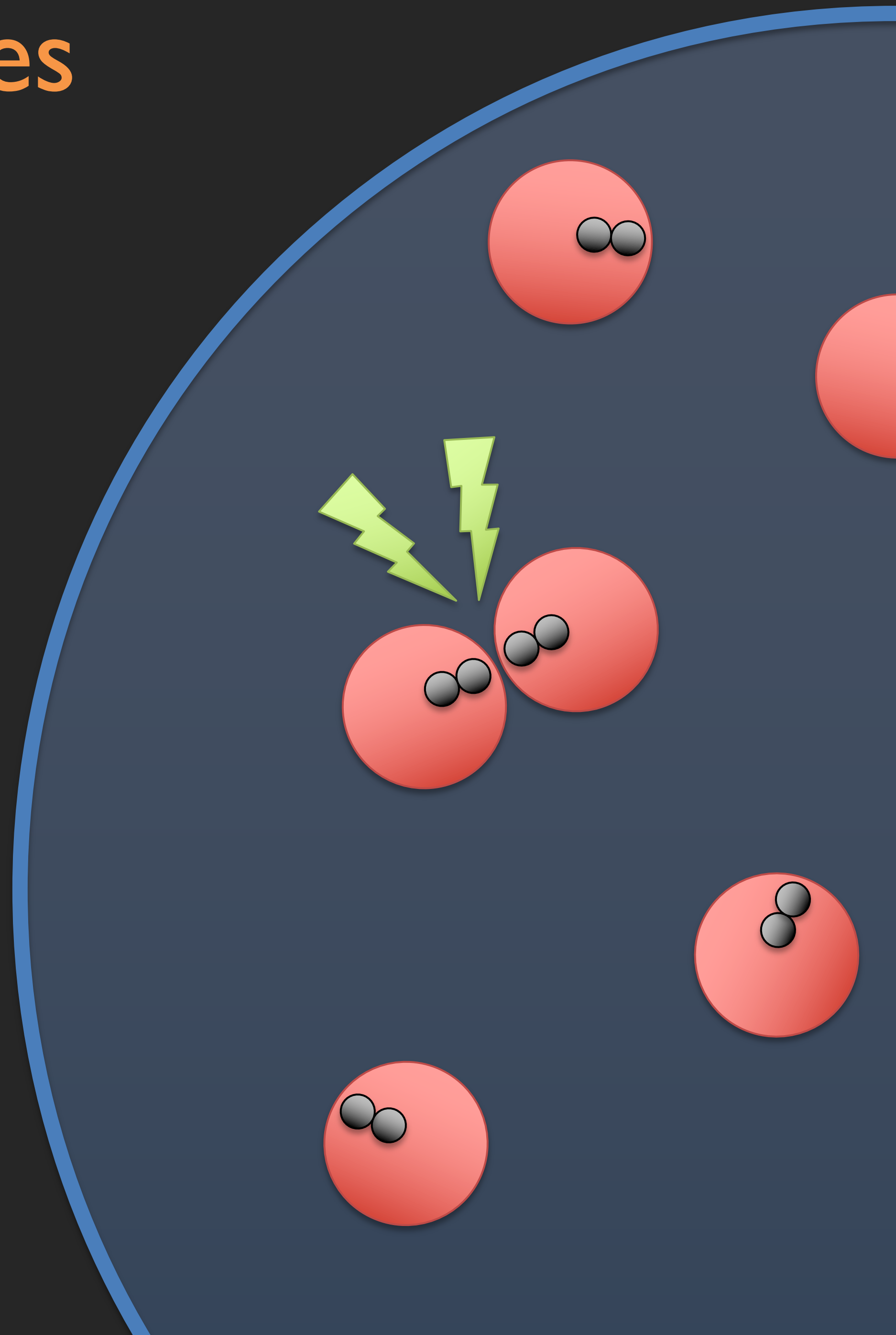
4. show evidence of being able to translate these understandings into the practical methodology of modelling

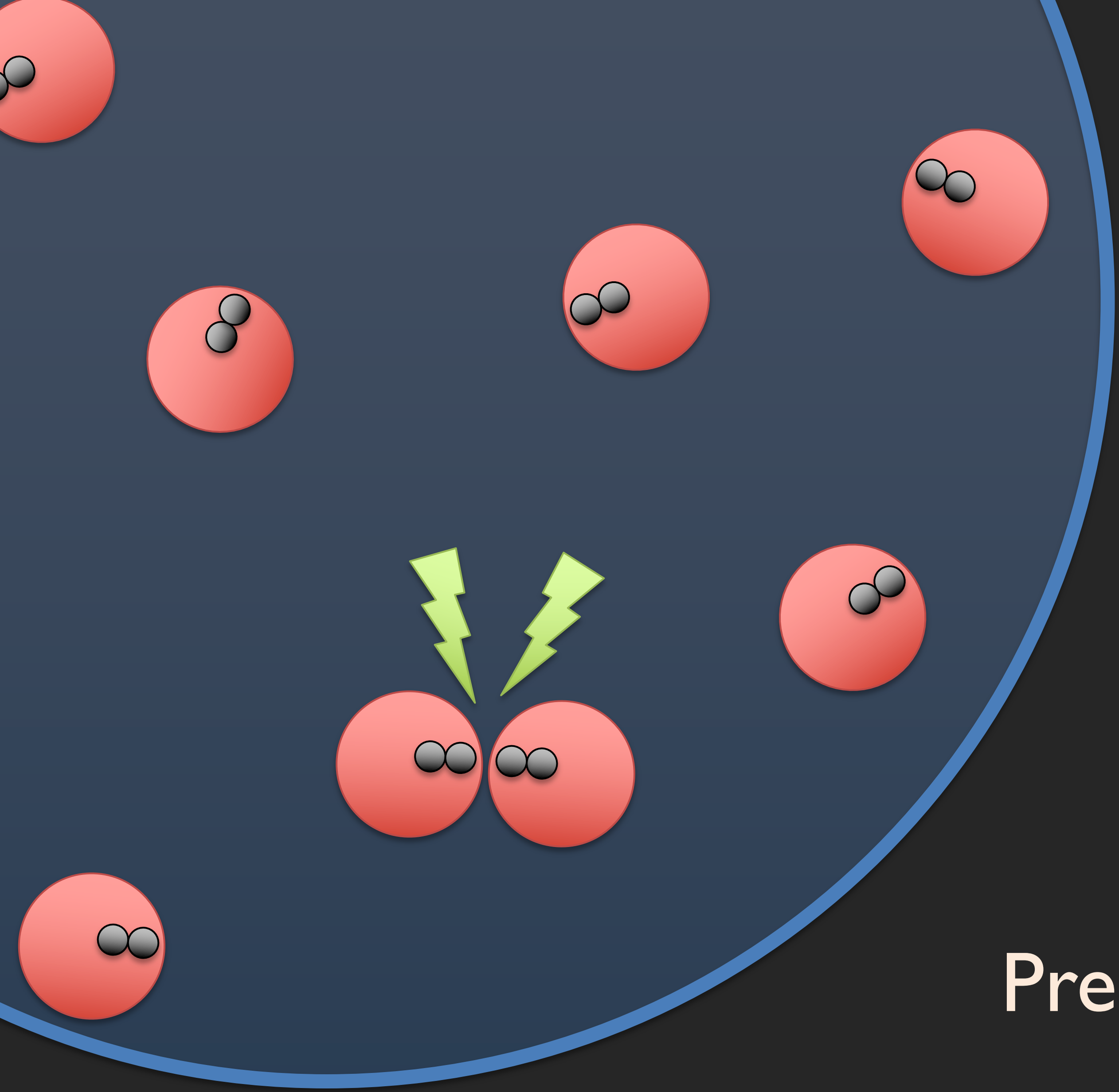
Session Objectives

1. Review the goals in presenting results

2. Be aware of the risks involved in assessing model outcomes

3. Look over some examples of the presentation of results





LECTURE 7

Presenting Results

Managing your results

- Preserve **raw and intermediate** forms of data
- **Document** all data cleaning and processing steps
  - maintaining a Jupyter notebook is an easy way to do this!
- **Script** your data analysis
  - manually wrangling an Excel sheet will introduce errors
- **Archive your data** in a DOI-issuing repository that provides citable URLs for specific versions of your dataset
- Create **tidy, analysis-friendly data**
  - choose the right metrics from the beginning

OpenABM Website – Guides to Good Practice

Let’s look back to Pattern-based Modelling!

What is Pattern-Oriented Modelling?

Researchers can "use multiple patterns as indicators of a system's internal organisation"

- Railsback & Grimm, p 228

Using "patterns" to:

- design the model's structure
- develop and test theory for agent behaviour
- find good parameter values

In essence: making sure a system works across a range of different dimensions!

Defining Patterns

A pattern is "**anything beyond random variation**"

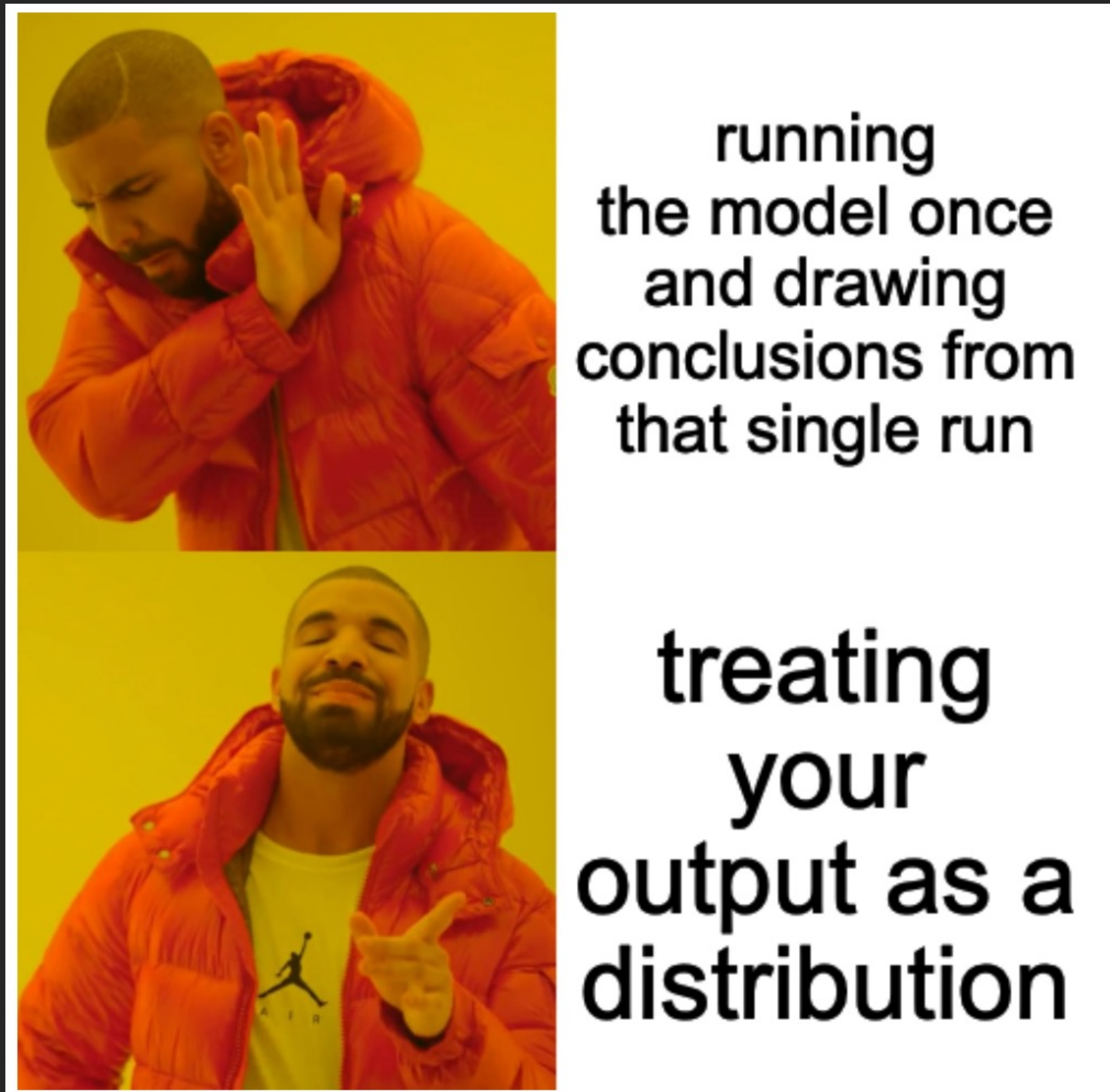
R&G, p 228

"**stylised facts** are broad…generalisation of empirical observations and **describe essential characteristics** of a phenomenon"

R&G citing Heine et al. 2007, who are in turn citing Kaldor 1961

- strong vs weak patterns - e.g. distinct cut-off versus parameters staying within a range as system responds in characteristic ways
- The hard part: **definition** of criteria for when a pattern is matched!

SO HOW DO WE PRESENT THESE PATTERNS?


running the model once and drawing conclusions from that single run
treating your output as a distribution

Make sure you are aware of:

- Which **metrics** you're testing and how you're going to compare them between scenarios.
- **Variation among runs** with the same parameters.
  - The mean might be a bad metric to use!
  - Look for **the distribution that characterises this data** –what does that mean for summary statistics??
- **Initialisation** and its impact.
  - How important is the initial set-up, and how much does it change?
  - Warm-up time!
- **Parameter sensitivity**.
  - Our friend the sensitivity test!

Choosing your metrics

This will depend on the context of your model, and you need to read up on the specific discipline to make sure it's going to be accepted by the experts!

- What metrics are other people who write about this using? Look in journal papers to understand the current best practices!
- What data are you trying to use to validate your model? What's available?
- Which statistical tests will allow you to assess how likely it is that the dynamic you're claiming you observe is, in fact, representative?

Imagine that I tell you I rolled a single die multiple times and I got the following numbers:

5, 2, 4, 1, 1, 6

15, 2, 4, 1, 11, 6 (We can pretty comfortably reject the hypothesis that these came from a 6-sided die!)

A worked example of "Confidence"

Statistical **confidence** is a measure of how **often** (the confidence **level**) you expect your estimate to fall within the **confidence interval**, if you redo the test multiple times.

For example: if you roll a die 100 times, how many 6s would you expect to get?

Calculating confidence intervals

If you roll a die 100 times, how many 6s would you expect?

You will need:

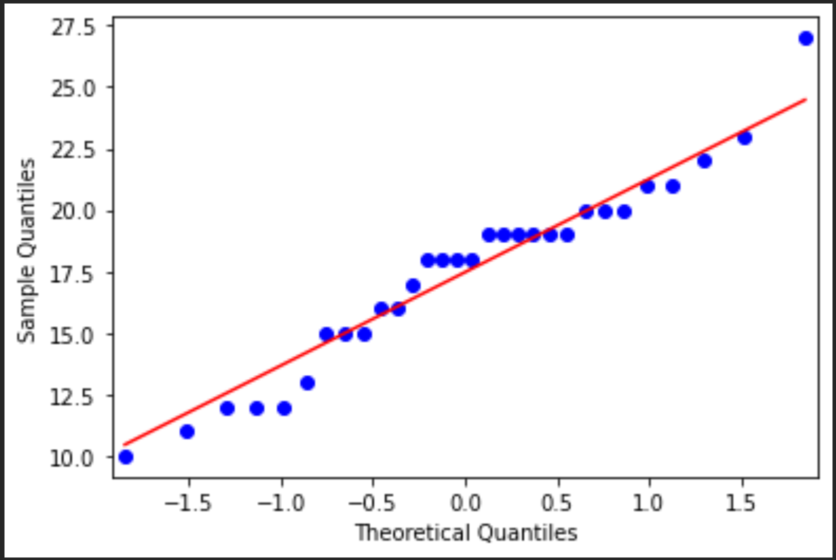
- The **point estimate** of your data (e.g. the **mean**).
- The **standard deviation** of your data.
- The **size** of your sample (number of points in your data).
- The **critical values** of your **test statistic.**
  - **Alpha** (level of confidence)
  - One- or two-tailed?
  - Look up critical value!

One example run:

4, 4, 4, **6**, 2, 3, 3, 4, 2, 4, 2, 2, 4, 4, 3, 1, 3, **6**, **6**, 2, 5, 2, **6**, 2, 1, 2, 5, **6**, 5, 1, **6**, 1, 1, 4, **6**, 5, 5, 3, 2, **6**, 2, 4, 1, 5, 4, 5, **6**, 3, 3, 3, 1, 2, 2, 2, 5, 1, 3, 2, 3, 1, 4, 3, **6,** 2, 2, 2, 2, 3, **6**, 2, **6**, 3, 2, 2, **6**, 5, 3, **6**, 2, 5, 2, 2, **6**, 1, **6**, 4, 4, 1, **6**, 1, **6**, 2, **6**, 3, 3, 3, 3, 5, 1, 1

Calculating confidence intervals

If you roll a die 100 times, how many 6s would you expect?



- Mean: **17.467**
- Standard deviation: **3.792**
- Sample size: **30**
- Critical values
  - Alpha: **.05**
  - **Two-tailed**
  - Large sample size (n>=30) and data is normally distributed (see QQ plot!) → use the Z score! **1.64**

(90% confidence level)

30 repetitions of this experiment:

13, 19, 23, 18, 20, 12, 15, 21, 19, 12, 19, 16, 19, 27, 17, 19, 11, 15, 21, 20, 18, 18, 19, 20, 22, 12, 15, 18, 16, 10

$$CI = \bar{X} \pm Z^* \frac{\sigma}{\sqrt{n}}$$

- CI = the confidence interval
- $\bar{X}$ = the population mean
- Z^* = the critical value of the *z* distribution
- σ = the population standard deviation
- √n = the square root of the population size

$$CI = 17.467 \pm 1.64 \frac{3.792}{\sqrt{30}}$$

$$CI = 17.467 \pm 1.135$$

If we repeat this experiment 100 times, we expect that 90 times, the number of sixes we find will be between 16.3 and 18.6.

Statistical tests: use the right language

Let's say you want to make a claim about a model. This claim claim, or hypothesis, posits that IF one thing happens/is changed, THEN another thing will happen/change.

You will need:

- **Some data**: measuring the effect you're studying.
- **A hypothesis**: describing the relation you expect to find in the data.
- Therefore, a **null hypothesis**: a statement that you can either **reject** or else **fail to reject**.

Through the statistical test, you can assess how likely the distribution-generating process you're imagining is to have generated the kind of data you've observed.

…WHAT?!

What does that mean!?

Mystery Distribution

...5, 6, 3, 3, 3, 1, 2, 2, 2, 5, 1, 3, 2...

does this data **^^^**
look as if it came out of the same
kind of machine as this data
vvv

Uniform Distribution
unif(1,6)

...3, 4, 2, 4, 2, 2, 4, 4, 3, 1, 3, 6, 6...

What does THAT mean!?

”how likely the **distribution-generating process** you’re imagining is to have **generated the kind of data** you’ve observed”

Examples:

→ The 6-sided versus 20-sided dice

→ (discrete) UNIFORM distributions of 1-6 vs 1-20

→ A weighted versus an unweighted coin

→ BERNOULLI distributions of heads/tails

→ A weighted versus an unweighted coin, flipped many times

→ BINOMIAL distributions of heads/tails

clicky

A RESOURCE I QUITE LIKE

Statistical tests: use the right language

Let's say you want to make a claim about the **firefly** model. You believe that the "delay" tactic helps the fireflies coordinate faster than the "advance" tactic.

You have:

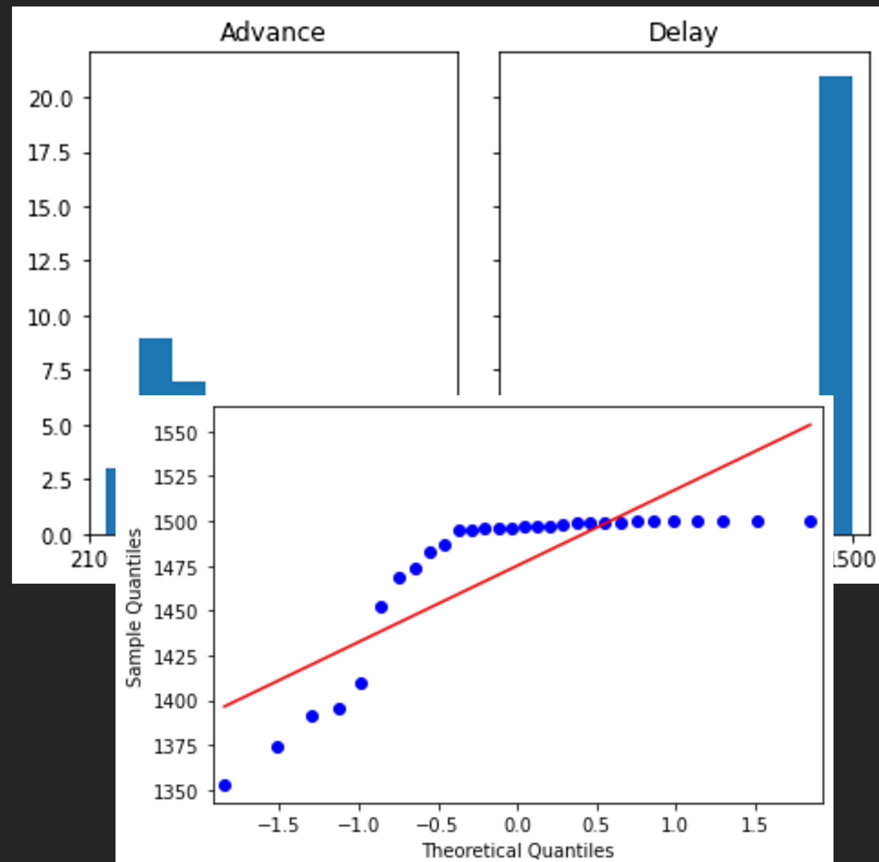
- **Some data**: the size of the biggest synchronized group of fireflies at the end of each run of the simulation.
- **A hypothesis**: the simulation finishes sooner if the fireflies are using the delay tactic.
- **A null hypothesis**: there is no meaningful difference between the results of these two parameter sets – the observations came from the same population

You're planning to show that in fact, the difference between the groups **IS** significant – by presenting data that allows you to **reject the null hypothesis**

Advance	Delay
237	1500
226	1500
217	1497
224	1500
212	1374
214	1495
221	1499
255	1500
225	1498
235	1499
220	1496
221	1499
219	1500
220	1410
220	1391
228	1497
223	1353
226	1483
218	1499
220	1496
241	1487
224	1497
237	1452
236	1396
224	1500
233	1500
218	1474
220	1495
223	1469
216	1496

We have the size of the largest synchronized group at the end of the model – now what?

In a perfect world: compare the distributions



Hmm, not…the most normal! Let’s try something else.

Advance	Delay
237	1500
226	1500
217	1497
224	1500
212	1374
214	1495
221	1499
255	1500
225	1498
235	1499
220	1496
221	1499
219	1500
220	1410
220	1391
228	1497
223	1353
226	1483
218	1499
220	1496
241	1487
224	1497
237	1452
236	1396
224	1500
233	1500
218	1474
220	1495
223	1469
216	1496

Wilcoxon Rank Sum Test for Independent Samples :
if one observation is made at random from each population (call them x_0 and y_0), then the probability that $x_0 > y_0$ is the same as the probability that $x_0 < y_0$

Using python:

```
from scipy.stats import ranksums

advance = [237, 226, 217, 224, …
delay = [1500, 1500, 1497, 1500,…
ranksums(advance, delay)

>> RanksumsResult(statistic=-6.6529,
   pvalue=2.8719e-11)
```

That means that if we rolled the dice 10^11 times, we'd expect to see something this extreme happen fewer than 3 times.

Therefore, we can reject the null hypothesis with confidence: these data come from distinct, distinguishable populations.

MORE GENERAL ADVICE

So what can we advise?

Be on the lookout for

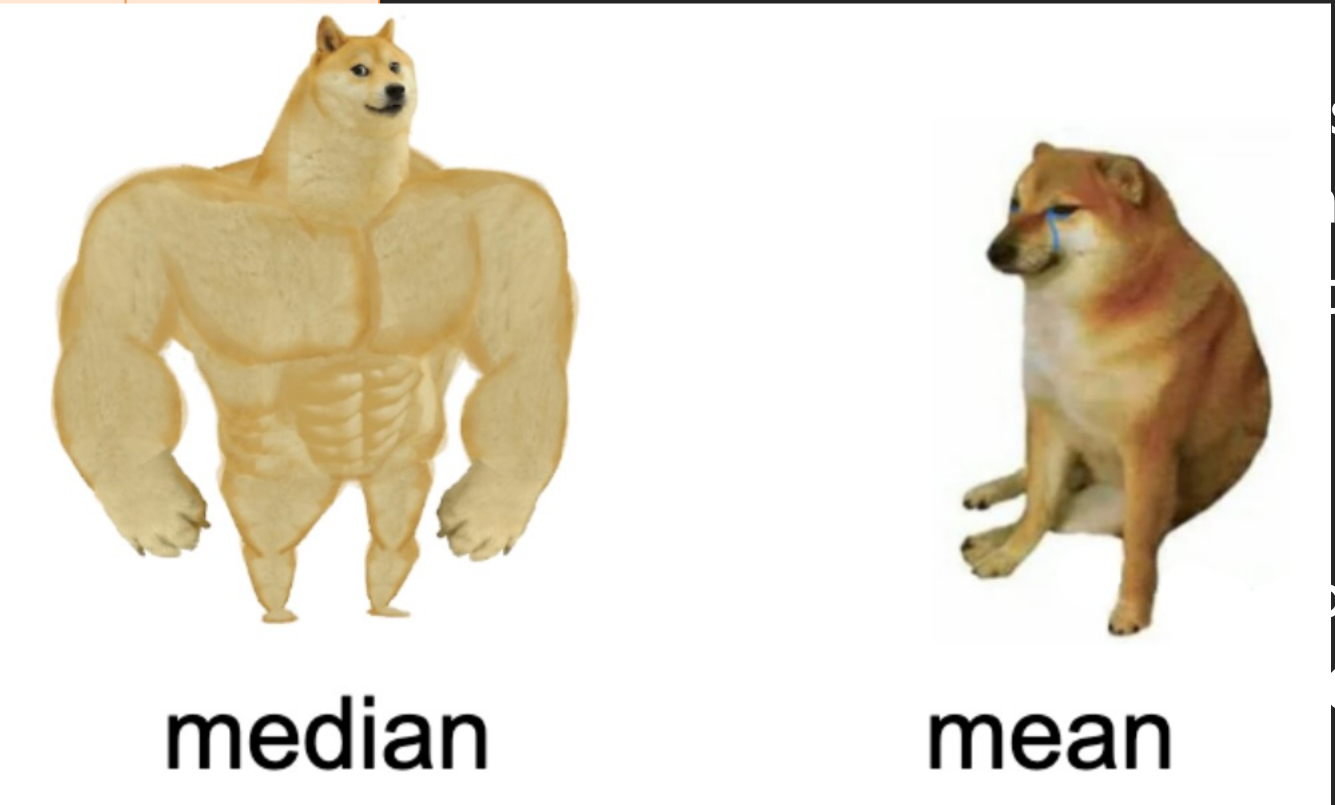
- Autocorrelation (spatially, temporally)
- Scale factors (e.g. the Modifiable Areal Unit Problem)
- Biases in sampling (in validation data or in your simulation)
- Built-in assumptions (e.g. starting from a uniform distribution of wealth versus normal or power-law distributed wealth)

Consider using

- Confidence intervals
- Error bars
- The median
- Inter-quartile ranges
- Logged values of your outputs (these may be normally distributed, which means you can use the other normal distribution tools on them!)

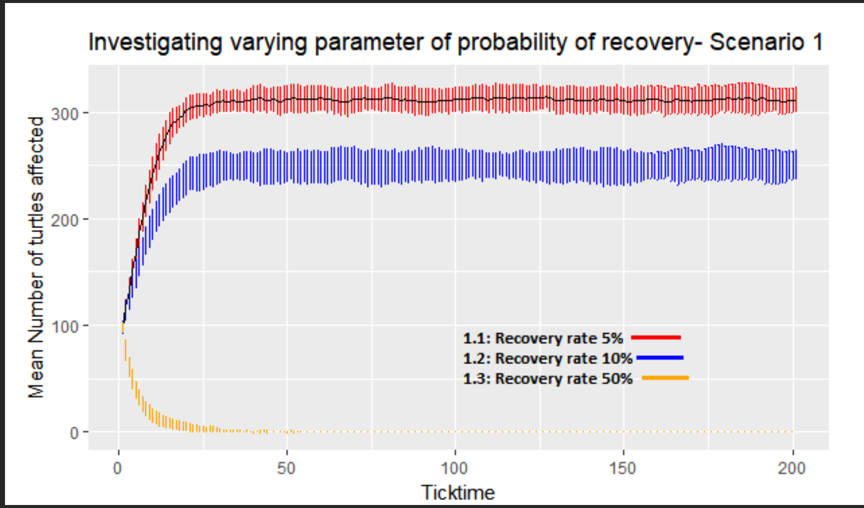
The peril of the mean

Run	Number of Cases
1	4
100	20984
Average	224.69



Variation

- Stochasticity is a big part of this field: respect it accordingly!
- Presenting both the median AND the mean is often useful.
- Plotting the "validation" value against the range of outcomes can be useful to assess how well it fits.
- Feeling uncertain? Look to the closest field expert!



(spatial autocorrelation, power laws, epidemic curve, Gini coefficient, etc.)

ABM-SPECIFIC WORRIES

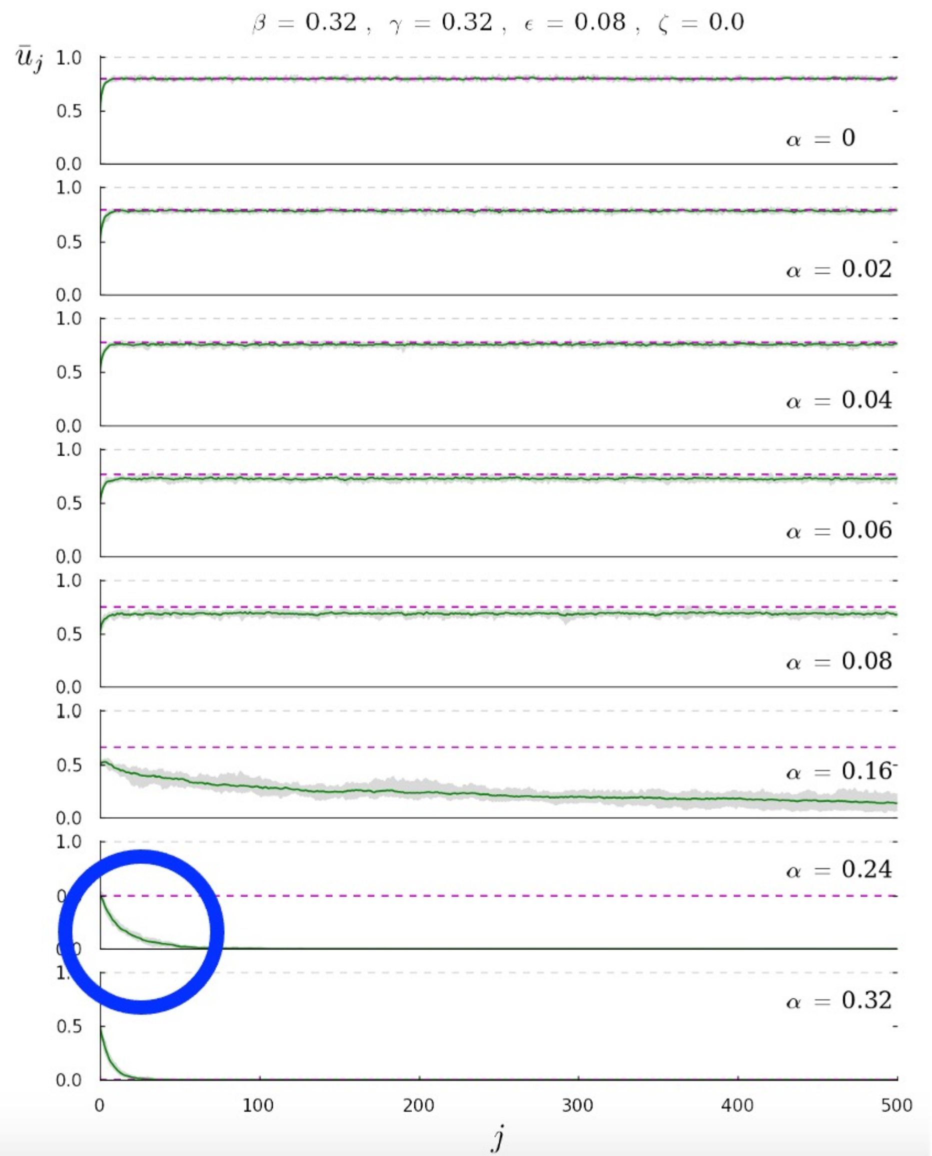
Example – Single Species B Models

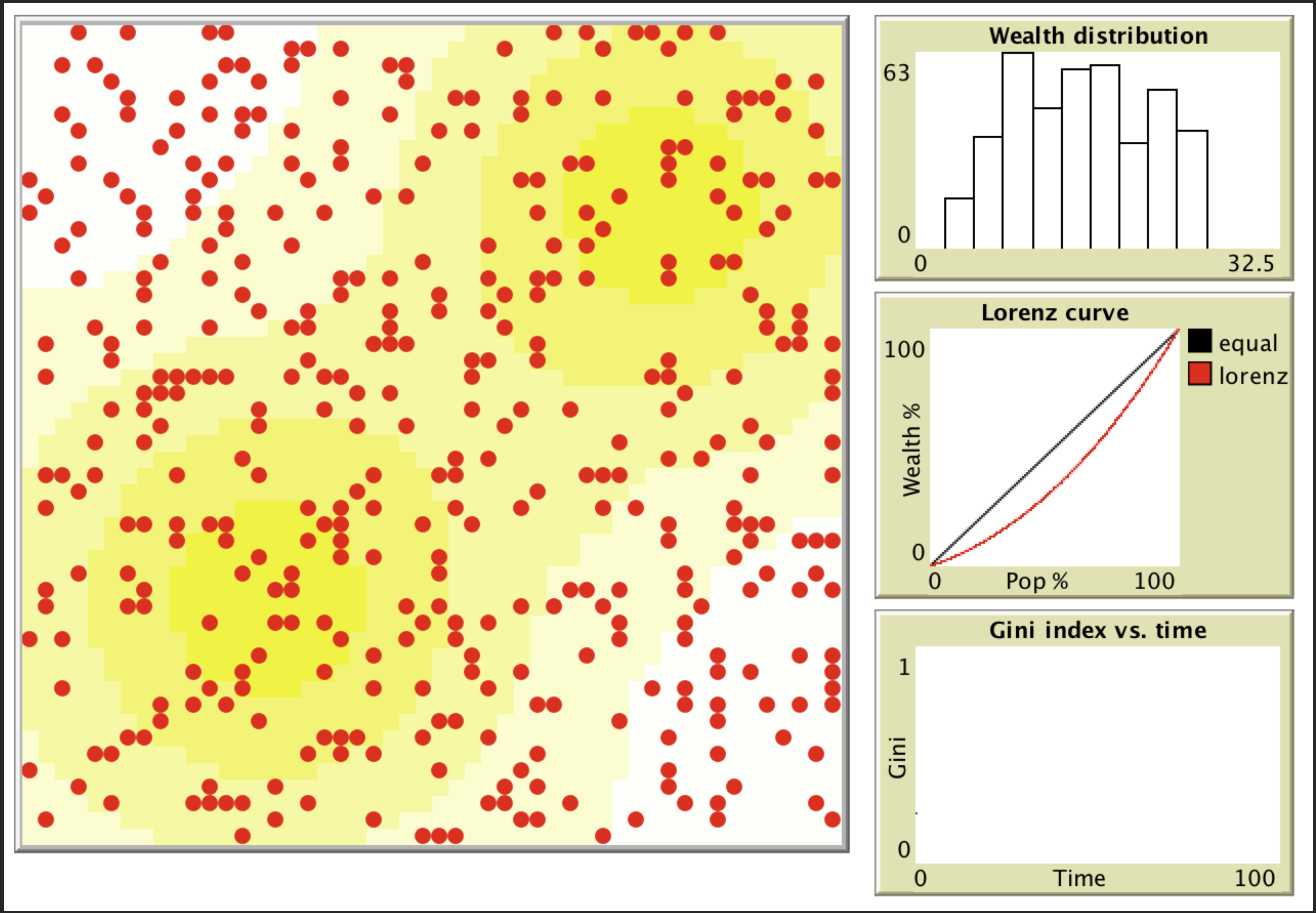
Approach II

Studying a single metric

Vary one variable, while holding others constant

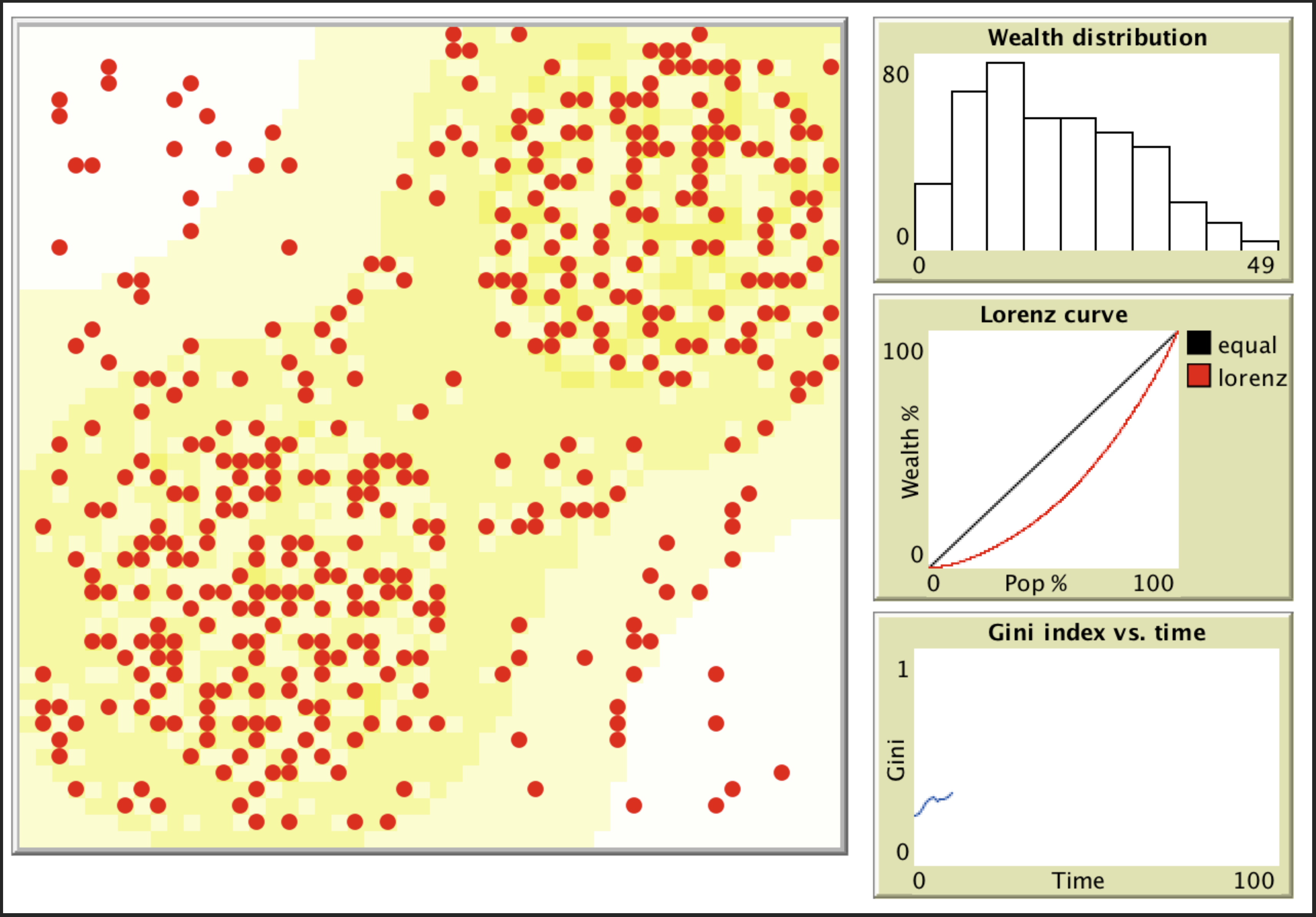
'Warm-Up' Period





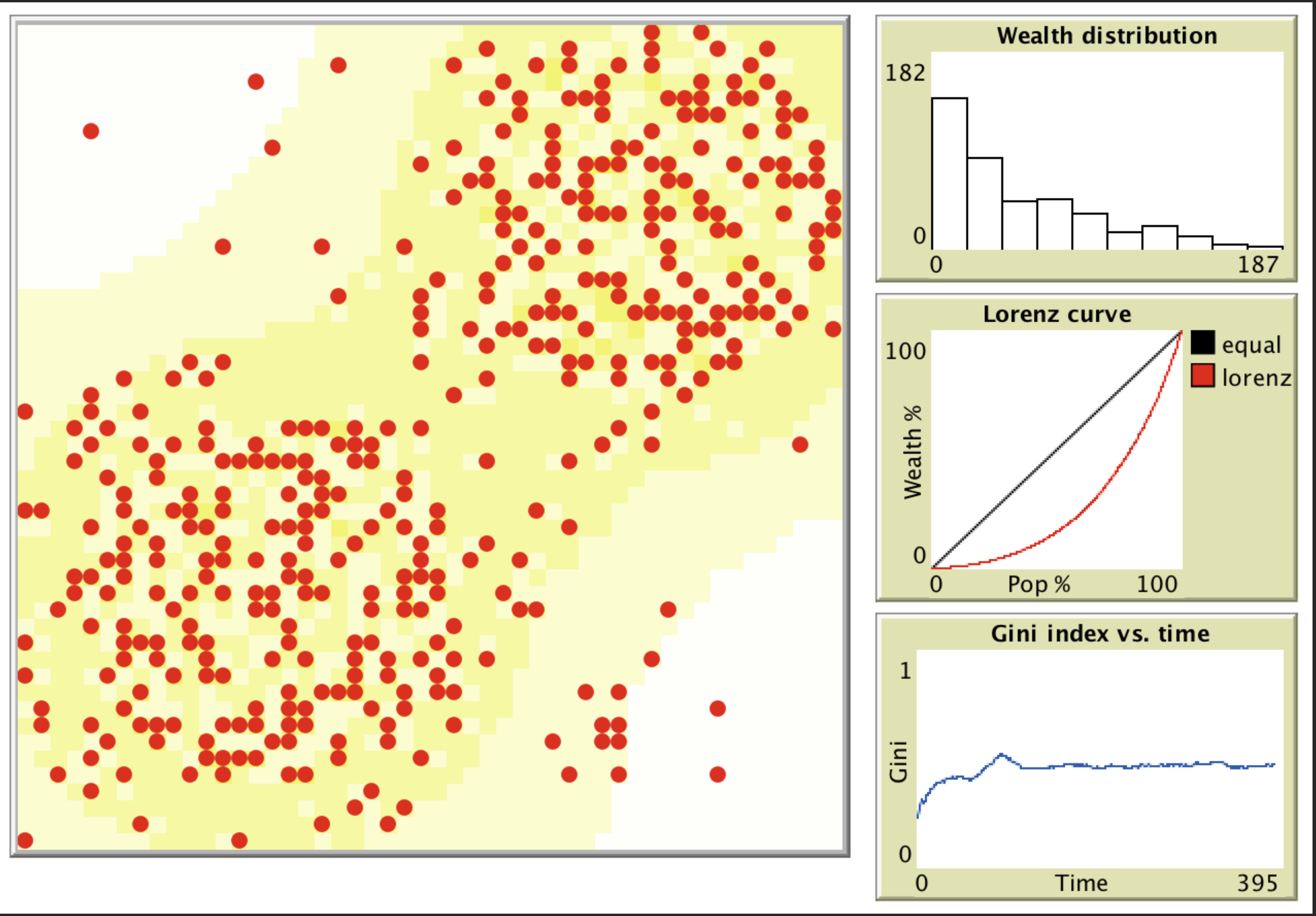
Wealth distribution
63
0
0
32.5
Lorenz curve
equal
lorenz
100
Wealth %
0
0
Pop %
100
Gini index vs. time
1
Gini
0
0
Time
100

Time 0

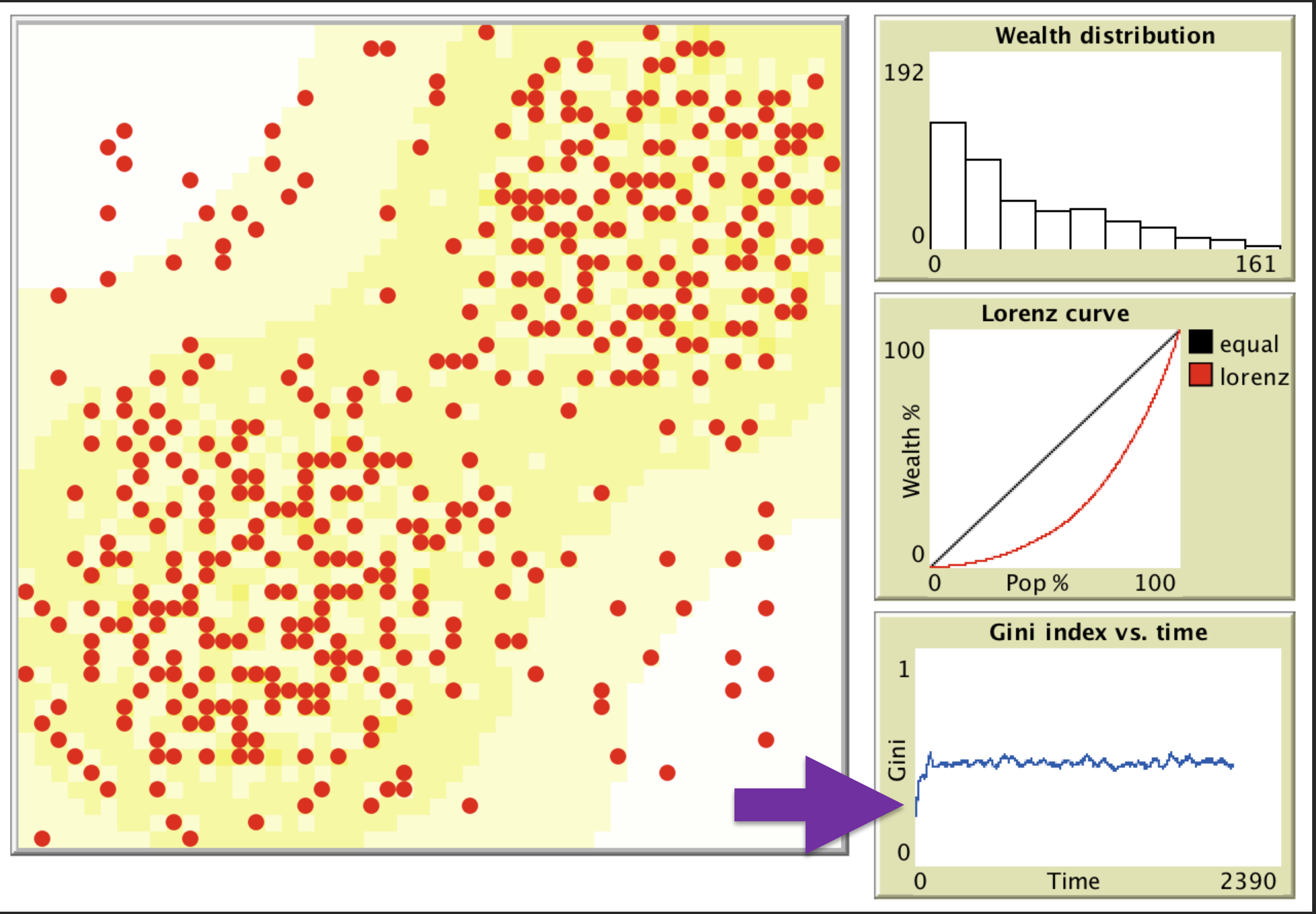


Wealth distribution
80
0
0
49
Lorenz curve
100
0
Wealth %
0
Pop %
100
equal
lorenz
Gini index vs. time
1
0
Gini
0
Time
100

Time 5


Wealth distribution
182
0
0
187
Lorenz curve
100
Wealth %
0
0
Pop %
100
equal
lorenz
Gini index vs. time
1
Gini
0
0
Time
395

Time 100



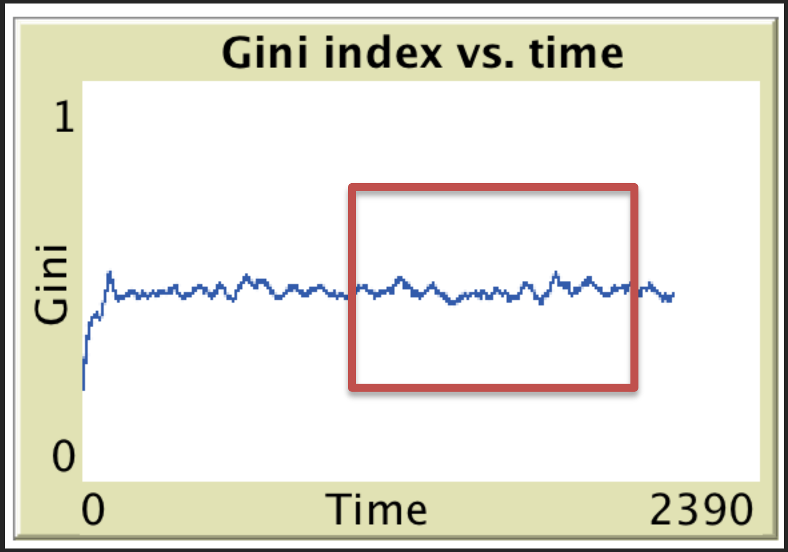
Wealth distribution
192
0
0
161
Lorenz curve
100
Wealth %
0
0
Pop %
100
equal
lorenz
Gini index vs. time
1
Gini
0
0
Time
2390

Time 1000

Best practice for warm-up period?

If you're measuring the range of outputs, you may want to exclude the warm-up period entirely

- Calculate over a **window** of the past results
- Keep in mind that different initial conditions may prompt different warm-up periods! This is something to explore



EXAMPLES IN THE WILD

	Private Car				Underground			
	Before	Percentage	After	Percentage	Before	Percentage	After	Percentage
Have car	101.69	46.22%	96.23	43.74%	30.31	13.77%	35.77	16.26%
No car	0	0%	0	0%	88	40%	88	40%
Total	101.69	46.22%	96.23	43.74%	118.31	53.77%	123.77	56.26%

Table 2 Changes in Commute Mode Choice of Residents Before and After the Completion of Elizabeth line

Items	Results
Degree of freedom	198
t	-7.7301
p-value	5.296e-13
Average of people with cars choosing underground before the completion of Elizabeth line (M_b)	30.31
Average of people with cars choosing underground after the completion of Elizabeth line (M_a)	35.77
The average increase of underground passengers with cars after the completion of Elizabeth line ($\Delta M=M_a-M_b$)	5.46

Table 3 Results of Two-sample T-test

	Private Car		Underground		Weight
	Before	After	Before	After	
Time	11.69 min	11.69 min	37.03 min	24.03 min	0.4
Cost	£0.57	£0.57	£2.40	£2.40	0.2
Comfort	0.84	0.91	0.33	0.30	0.4

Table 4 Changes of Factors for the Commute Mode Choices Before and After the Completion of Elizabeth line

Work of Chengru Deng

…the driving time is less than one third of the total time of underground before the completion of Elizabeth line. After its completion, the driving time remains unchanged, **but the travel time in the underground is shortened by 65%**, from 20 minutes to 7 minutes, which greatly improves the score of time for underground.

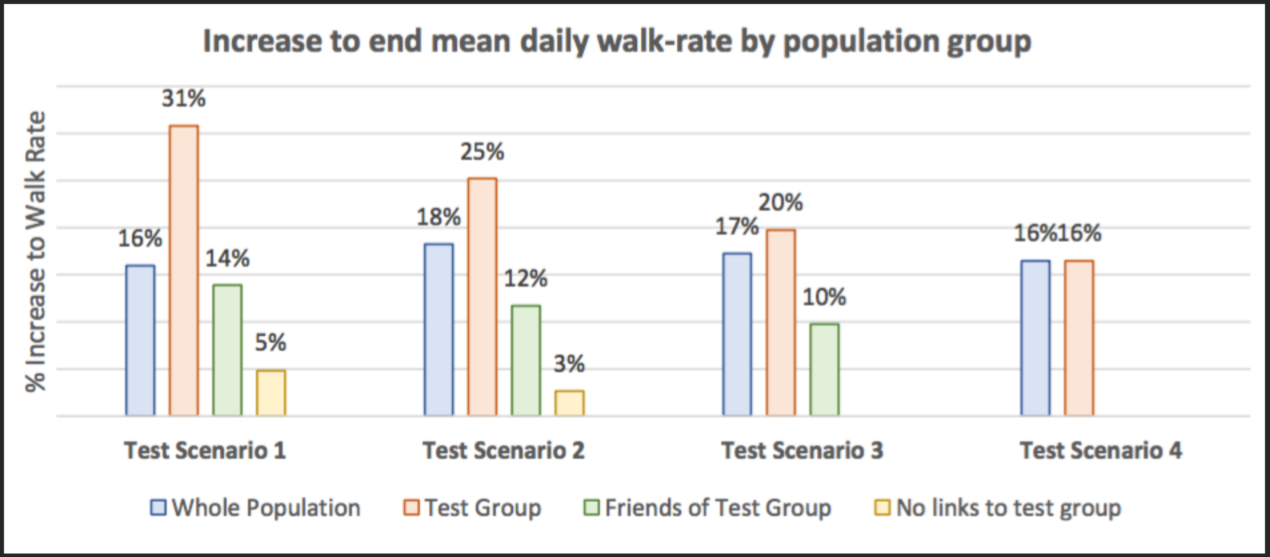
For cost, the cost of taking underground is nearly quadruple as that of driving private cars. After the completion of Elizabeth, the cost of neither commute modes is changed.

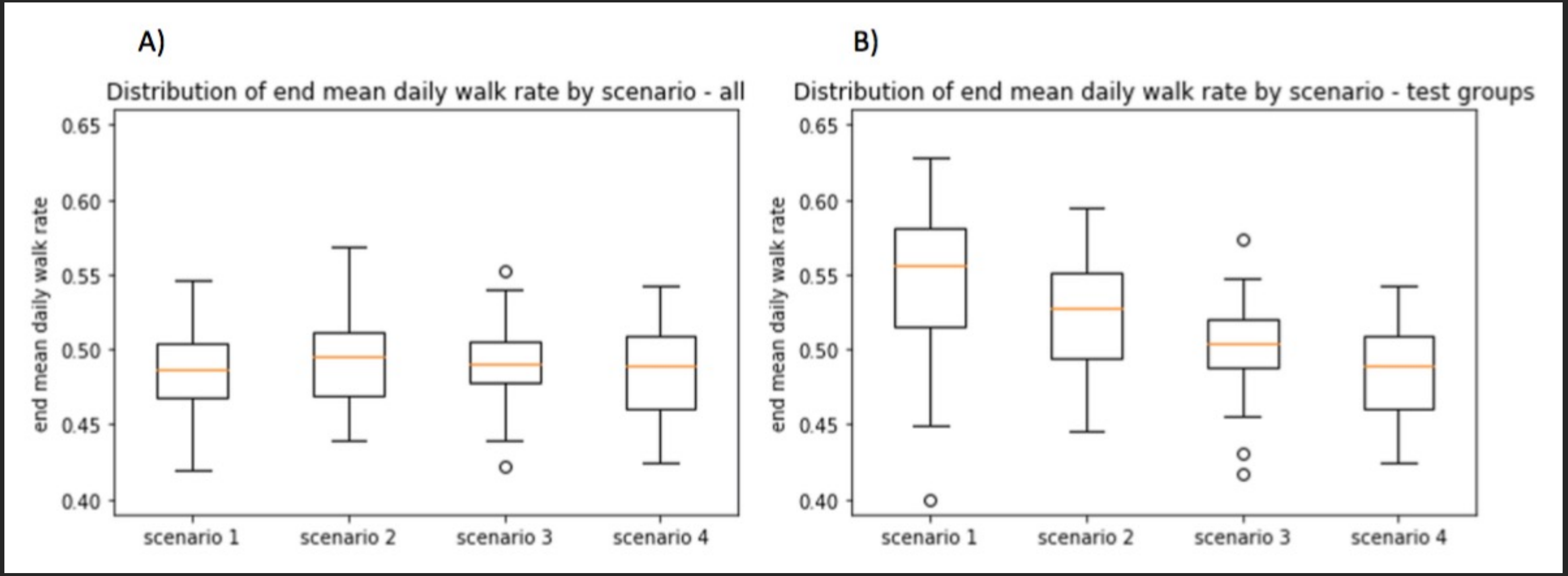
For comfort level, at first, the comfort level of private cars is more than twice as that of underground. After the completion of Elizabeth, more people are attracted to take the underground rather than drive their cars, which **intensifies the crowdedness in the underground and relieves the traffic congestion. As a result, comfort level of private car use is increased, while that of underground is exacerbated**.

However, even though people give the same weight to time and comfort factors, **the declined travel time of underground still outweighs the worsened comfort level, which encourages more people to take the underground**.

In conclusion, although private car use still attracts much more people who possess cars due to its much better performance in terms of all three factors, **the Elizabeth line is still successful in encouraging more people to shift from private car use to underground by its much shortened travel time**.

Work of Chengru Deng





Work of Greg Slater

However, conclusions shouldn't be drawn without examining the variation in results between runs. It is clear from figure 5A that **although the mean for each scenario is different, the distribution of results sit across very similar ranges**. This similarity is confirmed by running comparison t-tests between the end mean walk-rates of different scenarios, for which there are no p-values below 0.05, suggesting that **the population level changes aren't substantially different between scenarios**. Because of the number of randomly assigned variables involved in the decision-making process, individual level behaviour can vary dramatically between runs. This means when group sizes are relatively small – in some of the scenarios they may be as low as 125 – there can be high variance in the walk-rate outcomes. This is clear in figure 5B, where variance is highest for scenario 1.

As well as the average-metabolism decreasing with higher initial-population, the variance of results also decreases

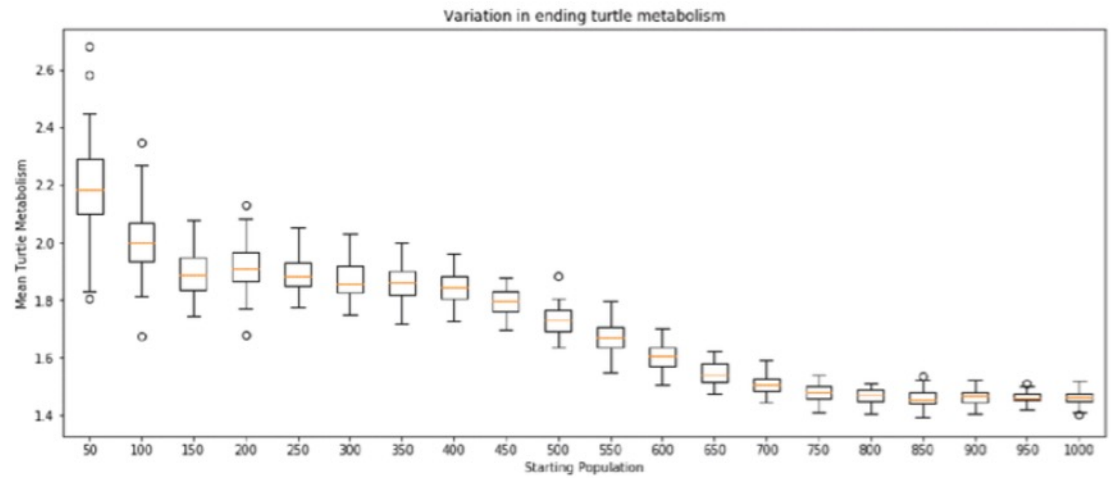


Figure 3

The variance in average-vision reduces initially but then stays relatively consistent suggesting the the relationship with initial-population isn't as pronounced as for metabolism.

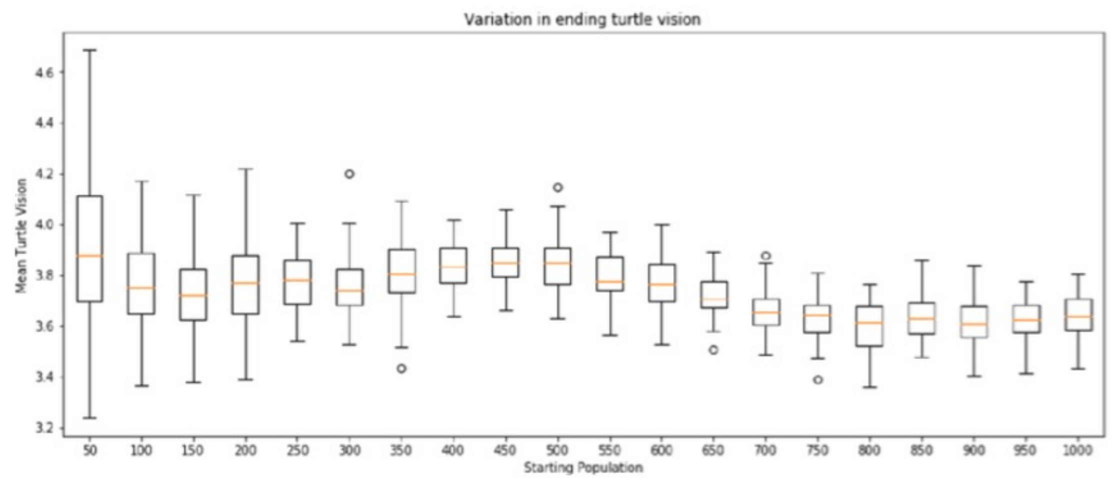
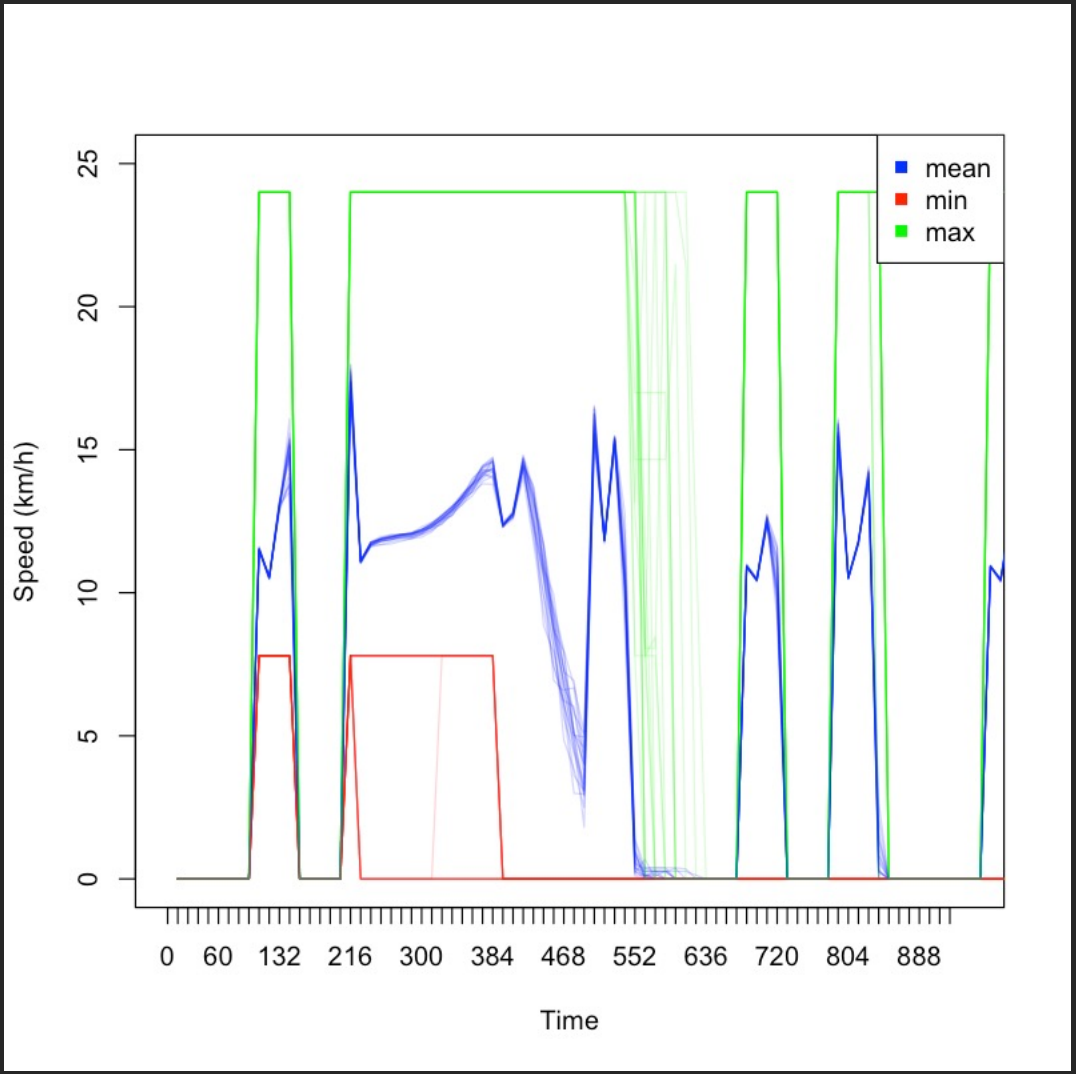
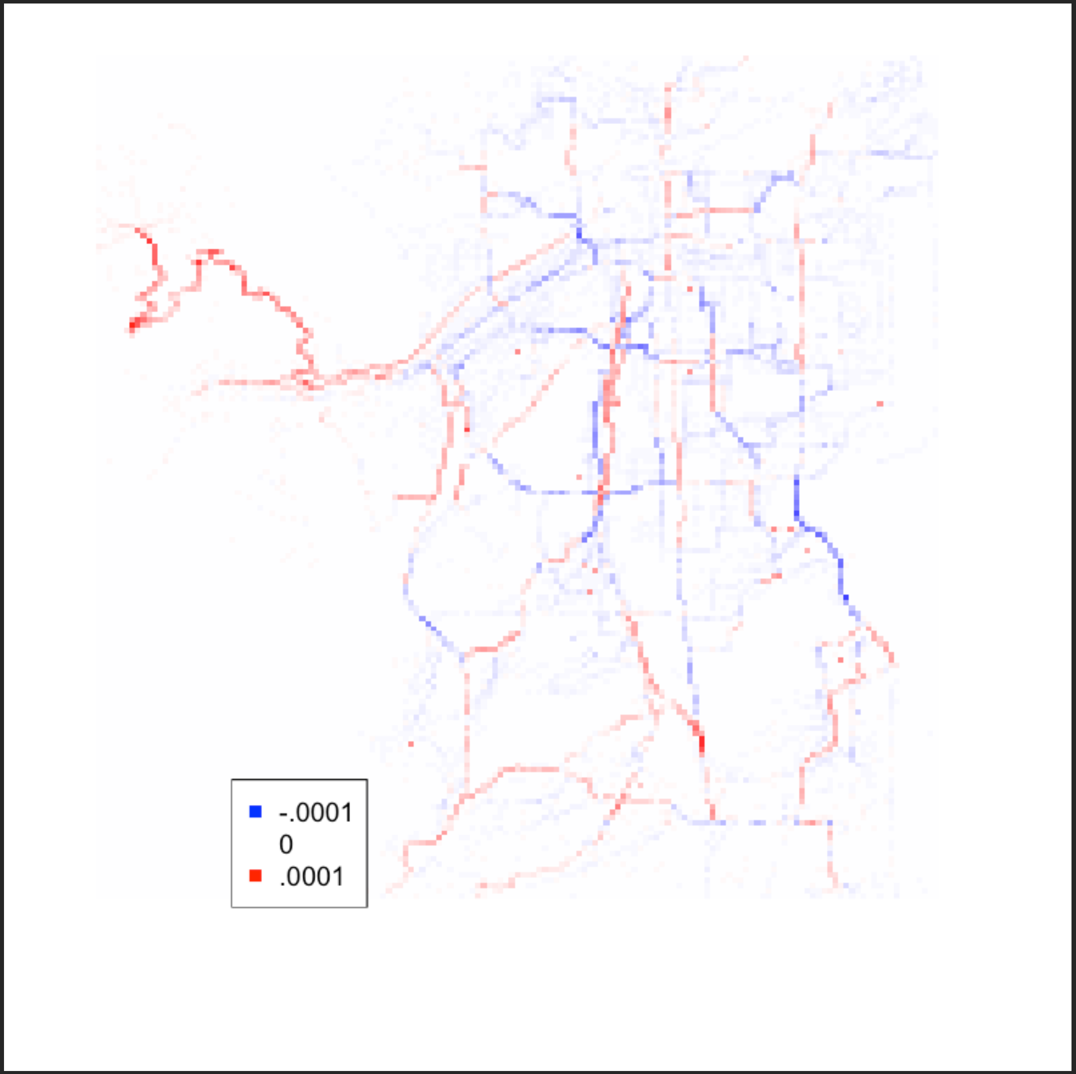
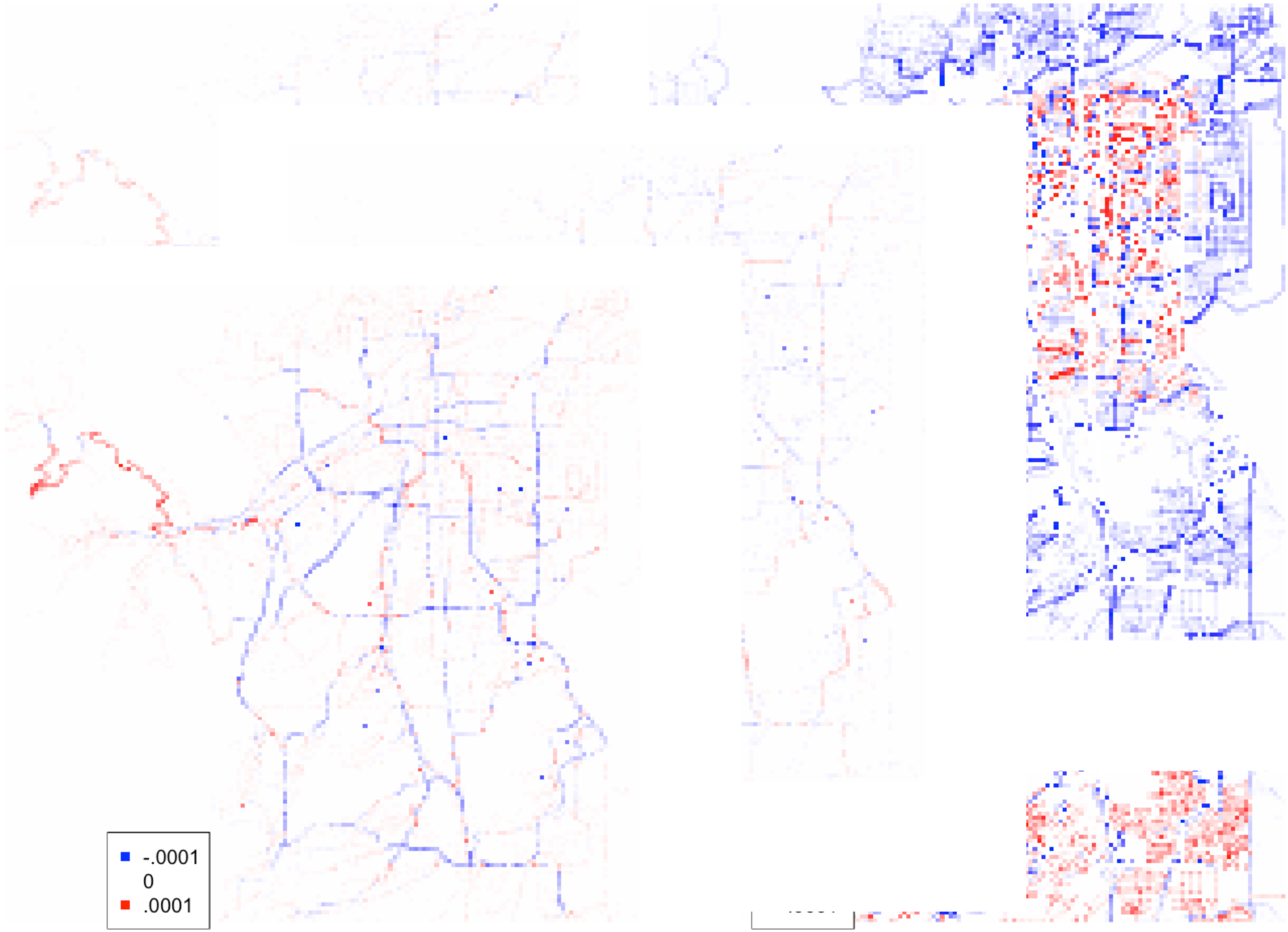


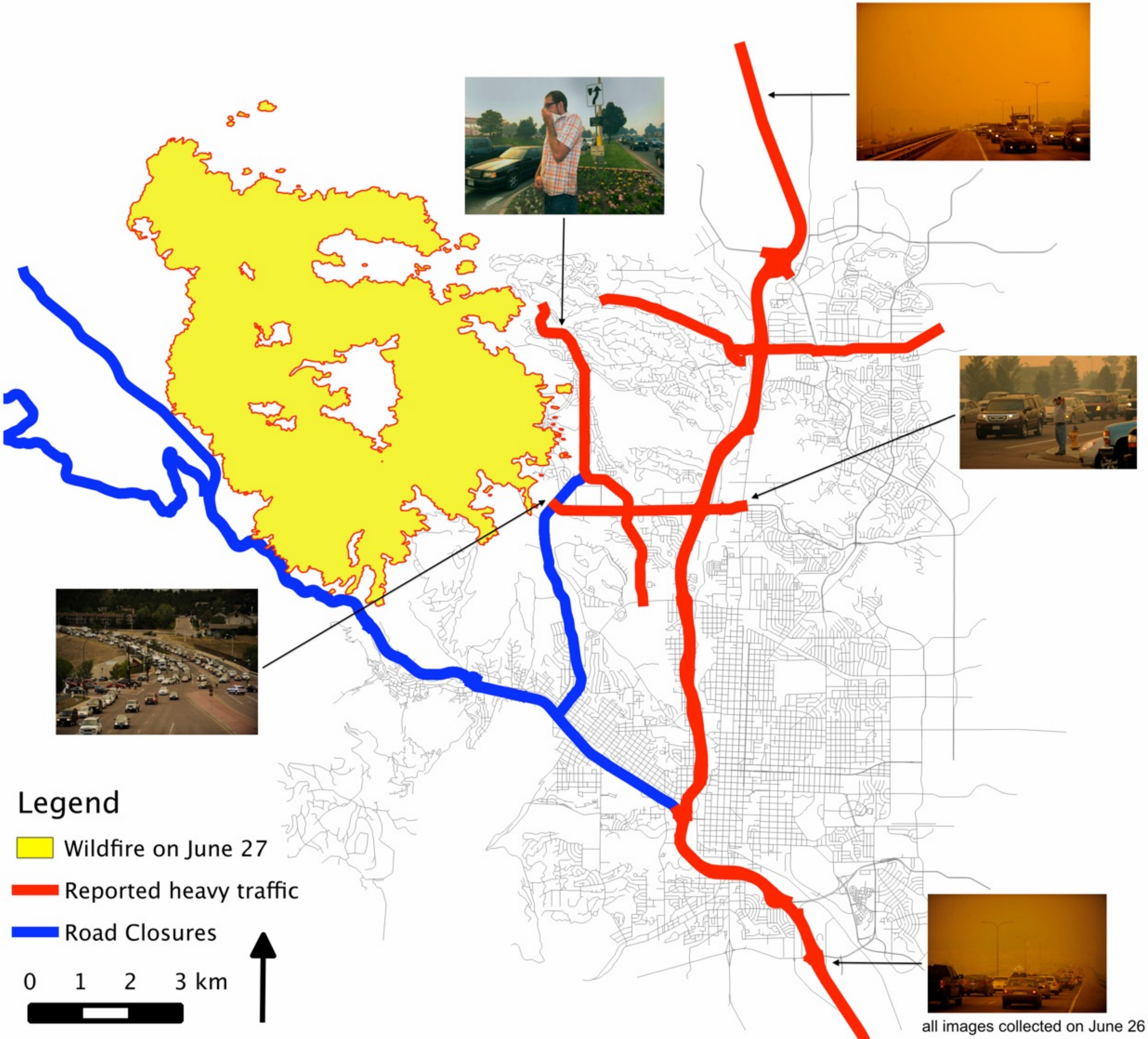
Figure 4

Work of Greg Slater


mean
min
max
Speed (km/h)
0
5
10
15
20
25
0
60
132
216
300
384
468
552
636
720
804
888
Time


-.0001
0
.0001


-.0001
0
.0001


Legend
Wildfire on June 27
Reported heavy traffic
Road Closures
0 1 2 3 km
all images collected on June 26

TAKE A BREATHER!