My experience with the RSS Statisticians for Society scheme

Clement Lee

2023-03-10 (Fri)

Newcastle University

Outline

- Statisticians for Society
- Working with a charity and their data
- Spatial analysis
- Thoughts

Newcastle University

Statisticians for Society (S4S)

Newcastle University

Overview of scheme

- RSS matches charity organisations with statisticians
- A *pro bono* / volunteering scheme
- A committee overseeing the whole process

Newcastle University

Scoping

- Charity contacts S4S committee
- Scoper meets with charity to identify need & form of support
- Charity fills in the scoping form
  - Project title & summary
  - Skills / qualifications required
  - Number of volunteers & hours required
  - Nature of data

Newcastle University

Advertising

- Committee emails opportunities to those signed up
  - Information from the scoping form
- s4s@rss.org.uk
- Requirement: fellow of RSS

Newcastle University

My applications

- Unsuccessful the first time
- The advert I applied to the second time

> **Understanding if current boundaries are the most effective**
>
> **Estimated time: 12 hours**
>
> We are in need of a volunteer for an organisation that operates a platform where people can exchange items for free. It runs in a way where the 'seller' can advertise an item, and a 'buyer' can connect, and then after agreeing, can then collect the item for free.

- ≥ 1 statistician(s)

Newcastle University

After getting selected

- 3-way meeting: (on 202**2**-03-10)
  - S4S project manager (Amirah)
  - someone from Freegle (Edward)
  - the statistician (me)
- Talk about project proposal
  - Scope
  - Timeline (3 months)
  - Method
- Put what was discussed in the document
- Sent proposal to scoper to approve

Newcastle University

Working with Freegle and their data

Newcastle University

Online platform

- https://www.ilovefreegle.org
- Online dating for stuff
- Like eBay or Gumtree
  - no money involved
- Or donating stuff to charity shop
  - more targeted

Newcastle University

Steps

1. Offerer puts item on Freegle
2. Replier responds to the post
3. Offerer and replier agree time and place
4. Replier goes to offerer's to pick up item
5. Offerer marks item gone on Freegle

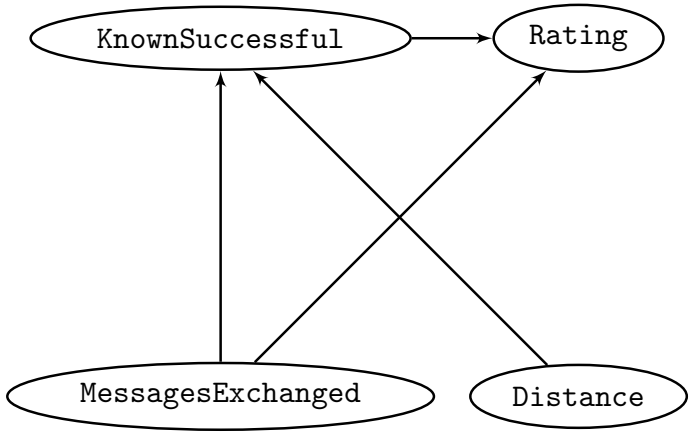
What the data look like

```
##    OfferID OfferLat  OfferLng OfferUID ReplyLat  ReplyLng ReplyUID
## 1 66473245 52.04057 -0.702386 10216160 52.02535 -0.801923 37937662
## 2 62004430 51.52760 -0.721791  2364225 51.53138 -0.720402   571023
## 3 59162925 50.80681 -1.876720  3467712 50.72900 -1.840794 36068373
## 4 54667917 54.07795 -2.840993  2060332 53.95756 -2.830094  3855713
## 5 62463673 51.58688 -1.795023 38607572 51.57926 -1.807035 33866461
## 6 58801500 54.55918 -2.496012   869151 54.79905 -2.642533   868189
```

```
##   MessagesExchanged KnownSuccessful PositiveRating NegativeRating
## 1                 6            TRUE          FALSE          FALSE
## 2                 1           FALSE          FALSE          FALSE
## 3                 4           FALSE          FALSE          FALSE
## 4                 6           FALSE          FALSE          FALSE
## 5                 8           FALSE          FALSE          FALSE
## 6                 6            TRUE          FALSE          FALSE
```

Newcastle University

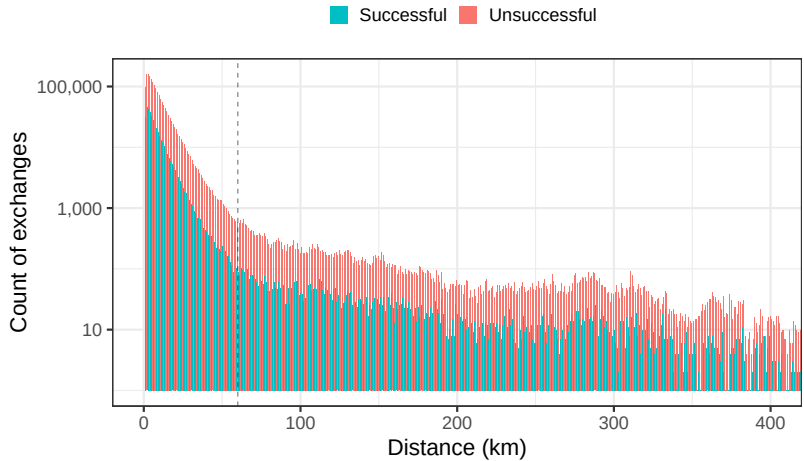
Causal diagram of variables



- What makes an exchange more likely to happen, & by how much?

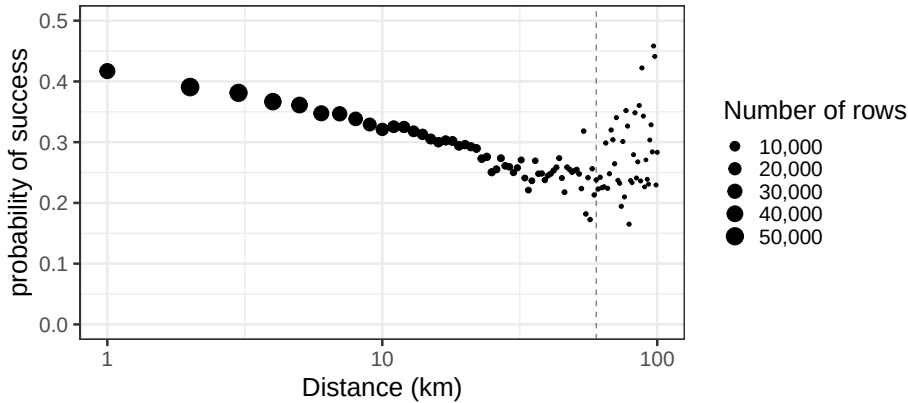
Newcastle University

Counts vs distance



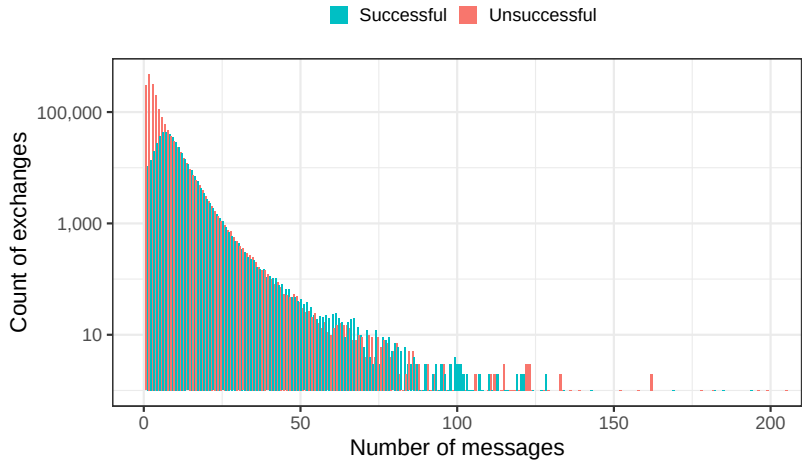
Newcastle University

Success probability vs distance



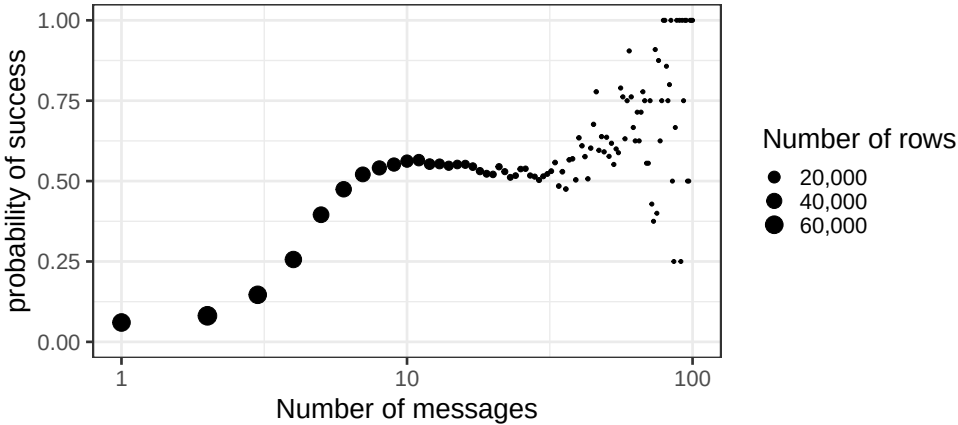
Newcastle University

Counts vs # messages



Newcastle University

Success probability vs # messages



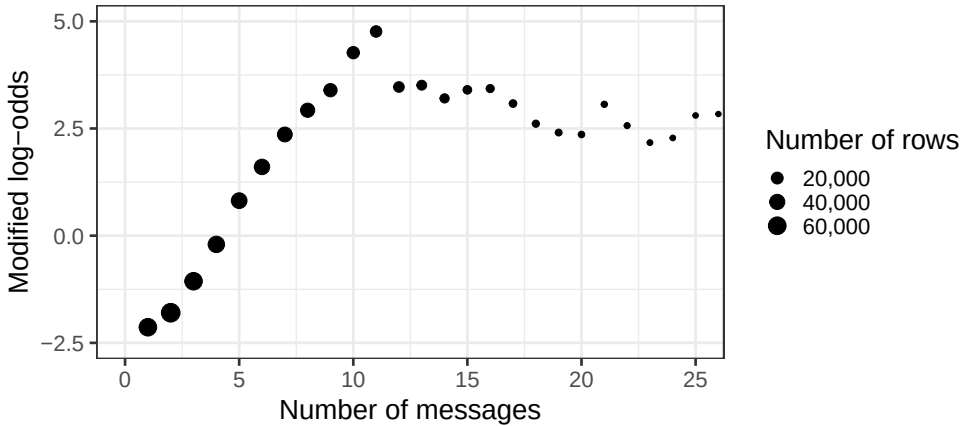
Newcastle University

Logistic regression?

- Probability plateaus around 0.57
- In real-life data, probability approaches 1 as the covariate (# messages) increases
- Replace $\log \frac{p}{1-p}$ by $\log \frac{p}{0.57-p}$ as linear combination of variables
- Call this the modified log odds

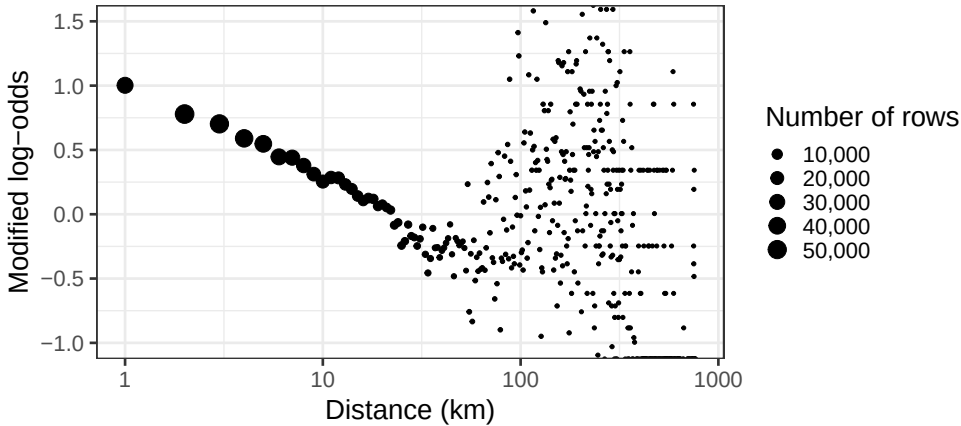
Newcastle University

Modified log-odds vs # messages



Newcastle University

Modified log-odds vs distance



Newcastle University

Modified logistic regression

Vanilla model

$$\log \frac{p_i}{1-p_i} = -0.618 + 0.0339 \times \text{Number of messages}_i - 0.148 \times \log(\text{Distance}_i)$$

Better:

$$\log \frac{p_i}{0.565-p_i} = -2.79 + 0.783 \times \text{Number of messages}_i - 0.261 \times \log(\text{Distance}_i)$$

Best:

$$\begin{aligned}\log \frac{p_i}{0.566-p_i} = &-3.14 + 0.885 \times \text{Number of messages}_i - 0.0653 \times \log(\text{Distance}_i) \\ &- 0.0562 \times \text{Number of messages}_i \times \log(\text{Distance}_i)\end{aligned}$$

Newcastle University

The mid-point call

- Me and S4S reviewers (and Amirah)
- I presented, they made comments
  - Interaction term
- An extension to the project deadline
  - Longer timeline
  - More hours

Newcastle University

Potential changes

- Nudging those who messaged too few times
- Minimal difference for those messaged more than 10 times
- Current effectiveness of the platform
- Concrete numbers help

Newcastle University

Writing up

- A final report for RSS & the charity
- Not a journal article
- Shared with others on the board of Freegle

Newcastle University

End of project call

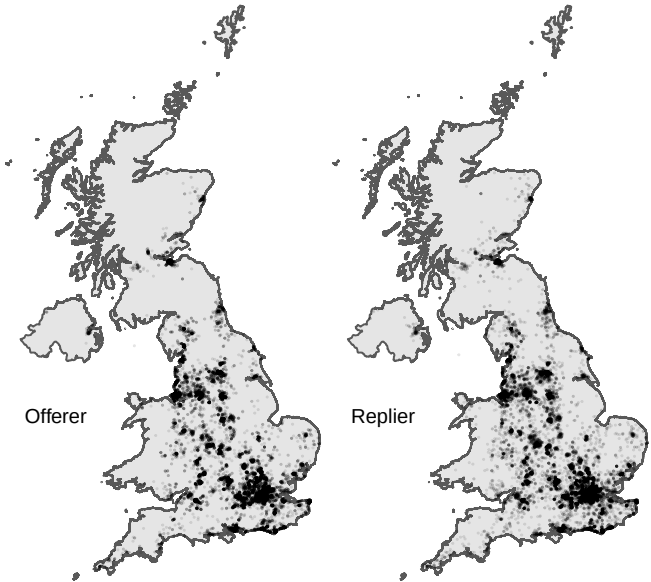
- More analysis & collaboration to follow, but “wrap-up” required for RSS
- 4-way meeting:
  - Edward
  - Me
  - Scoper
  - S4S project manager
- Scoper makes sure needs are met & everyone is happy
- Edward & I made recommendations

Newcastle University

Spatial analysis

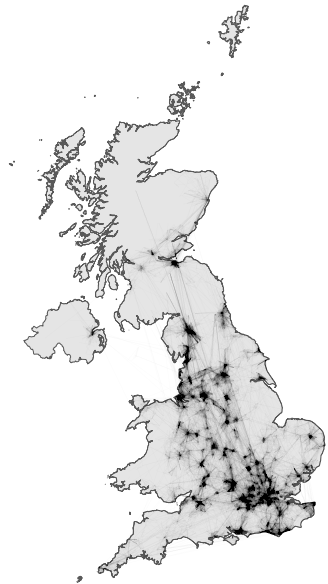
Newcastle University

Location of users



Newcastle University

Interaction network



Newcastle University

Questions

- Are the communities learned from the data similar to those created by the board / volunteers?
- How far is a user willing to travel for an item?
  - physical distance
  - adjustment by urban scaling factors, and/or
  - level of social deprivation

Newcastle University

Back to original project summary

Understanding if current boundaries are the most effective

Estimated time: 12 hours

We are in need of a volunteer for an organisation that operates a platform where people can exchange items for free. It runs in a way where the 'seller' can advertise an item, and a 'buyer' can connect, and then after agreeing, can then collect the item for free.

- I didn't answer this question
- MSc project!

Newcastle University

Some thoughts

Great match

- The charity knows what needs to be answered
- They know their data well
- Knowledge of statisticians complement what they have

Newcastle University

Tips

- Give concrete models and numbers
  - ~~high-level advice~~
- Coding
  - Data cleaning & visualisation
  - Reproducibility - one script to rule them all
- Meet in person
  - The human interaction

Newcastle University

Lastly

Case study:

- https://rss.org.uk/membership/volunteering-and-promoting/statisticians-for-society-initiative/case-studies/freegle/

Thank you

Newcastle University