SoDA 496 Introduction

Mike Burnham

1/11/23

Three Big Ideas

1. Data generating processes can be modeled.
2. There are intrinsic limits to what can be inferred from a model.
3. Your ideas are only as good as your marketing.

Why?

Why is this important?

Why is this important?

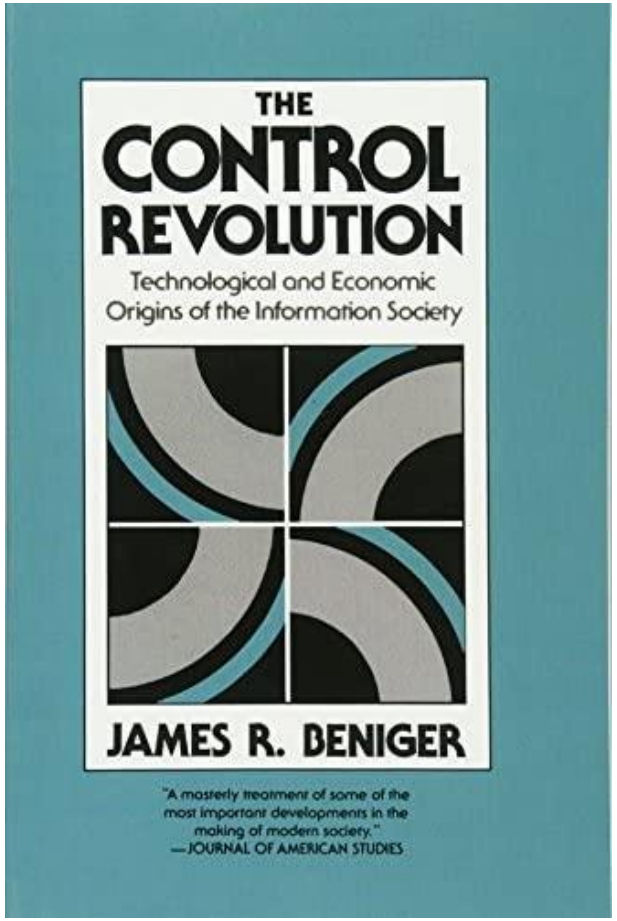
- Why does this matter for society?

Why is this important?

- Why does this matter for society?
- Why does this matter for science?

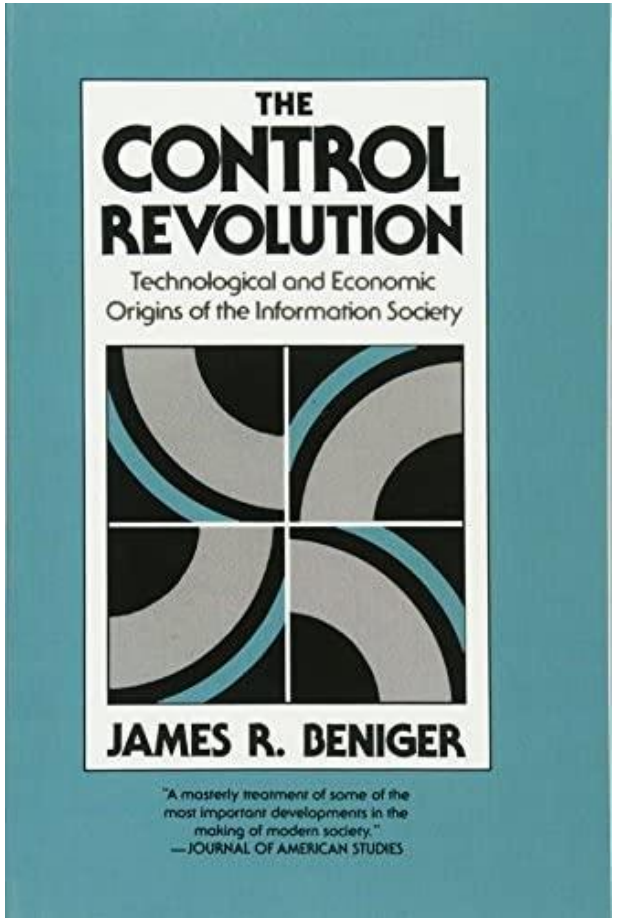
The Control Revolution

- The industrial revolution created a “crisis of control.”



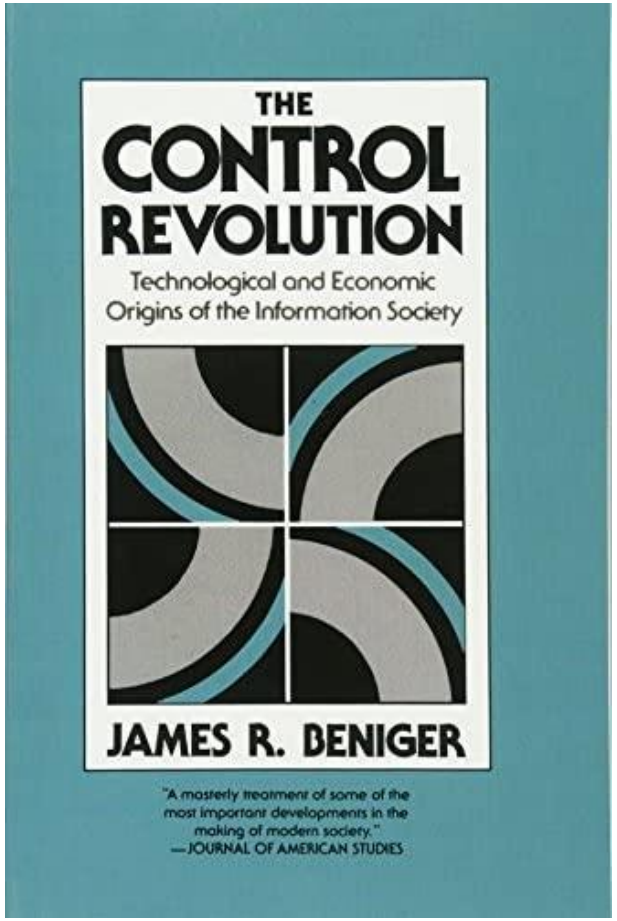
The Control Revolution

- The industrial revolution created a “crisis of control.”
- **Control:** Purposeful influence toward a predetermined goal.



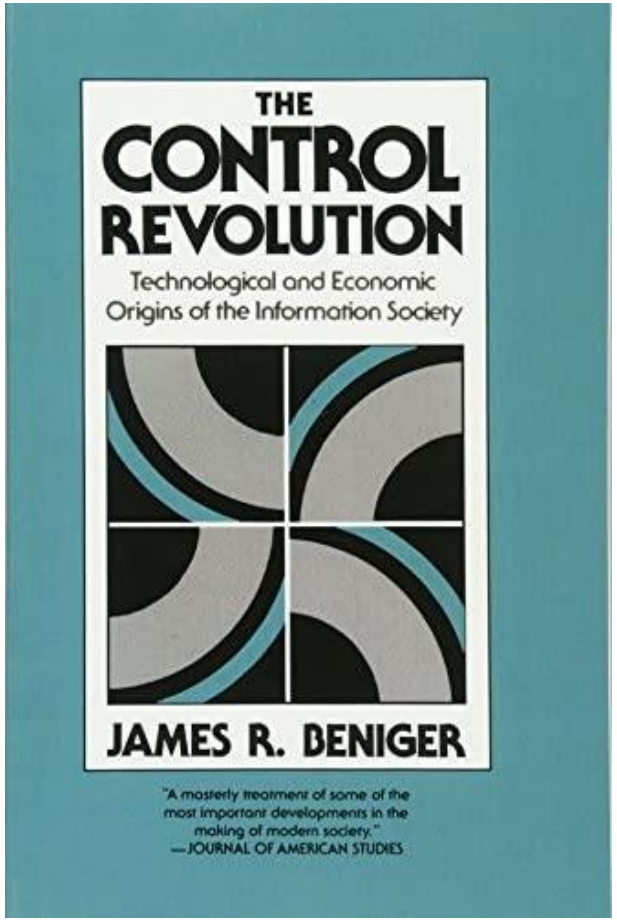
The Control Revolution

- The industrial revolution created a “crisis of control.”
- **Control:** Purposeful influence toward a predetermined goal.
- The increased speed and scale of society brought logistic and trust problems.



The Control Revolution

- The industrial revolution created a “crisis of control.”
- **Control:** Purposeful influence toward a predetermined goal.
- The increased speed and scale of society brought logistic and trust problems.
- A society’s ability to maintain control will be directly proportional to the development of information technologies.



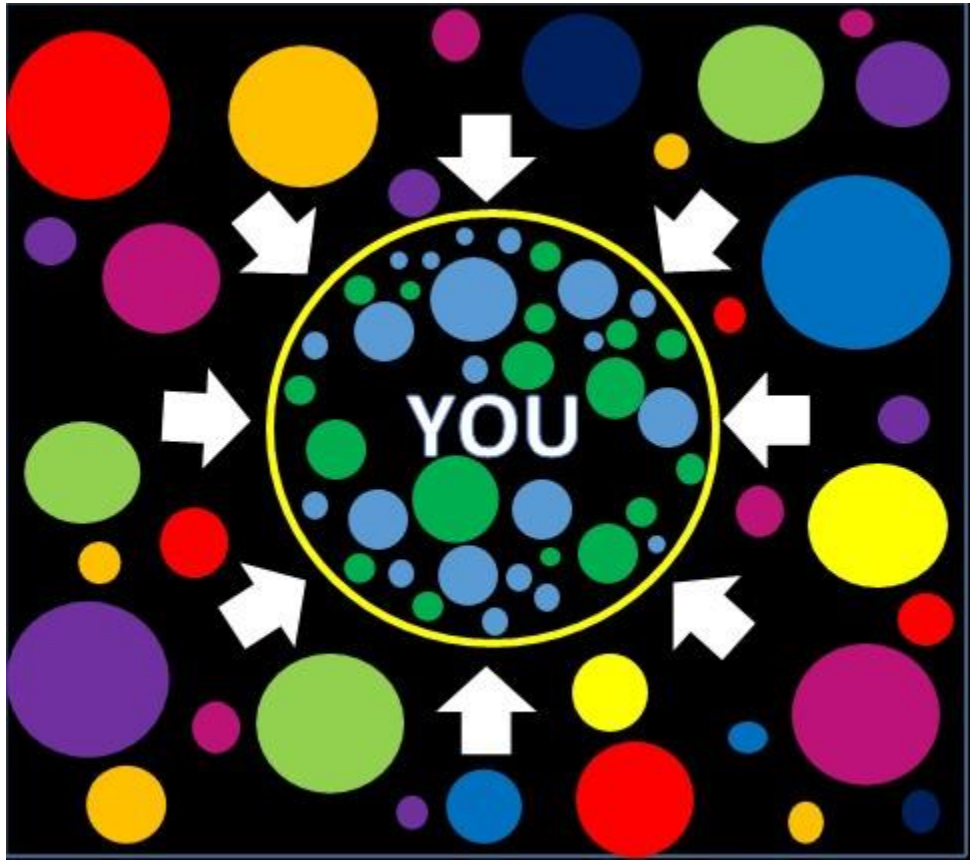
Scientific Advancement

- For most of human history our ability to collect and model data on human behavior was relatively limited

What is causing polarization?

A common answer:

A common answer:



However…

- Empirical evidence suggests that social media *increases* the diversity of media consumption (Scharkow, Mangold, Stier, & Breuer 2020).

However…

- Empirical evidence suggests that social media *increases* the diversity of media consumption (Scharkow, Mangold, Stier, & Breuer 2020).
- It isn't clear that echo chambers are polarizing. Some research suggests that it is, but other research suggests that it's exposure to different views that polarize (Bail et. al. 2018).

However…

- Empirical evidence suggests that social media *increases* the diversity of media consumption (Scharkow, Mangold, Stier, & Breuer 2020).
- It isn't clear that echo chambers are polarizing. Some research suggests that it is, but other research suggests that it's exposure to different views that polarize (Bail et. al. 2018).
- Further, data from around the world suggests that while social media use is a global phenomenon, polarization is not (Boxell, Gentzkow, & Shapiro 2021). Why?

However…

- Empirical evidence suggests that social media *increases* the diversity of media consumption (Scharkow, Mangold, Stier, & Breuer 2020).
- It isn't clear that echo chambers are polarizing. Some research suggests that it is, but other research suggests that it's exposure to different views that polarize (Bail et. al. 2018).
- Further, data from around the world suggests that while social media use is a global phenomenon, polarization is not (Boxell, Gentzkow, & Shapiro 2021). Why?
- This doesn't necessarily mean that social media has no impact on polarization, but it does imply that the popular model is wrong.

“Common Sense”

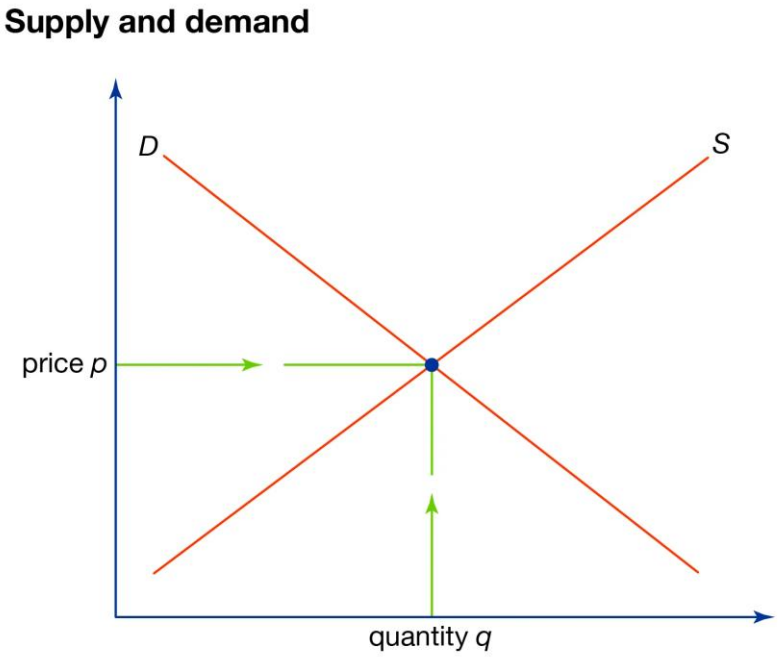
- Science has shown repeatedly that the reality of our physical and social worlds is often not what we expect. “Common sense” models often simply don’t reflect reality

“Common Sense”

- Science has shown repeatedly that the reality of our physical and social worlds is often not what we expect. “Common sense” models often simply don’t reflect reality
- Examples:
  - Echo chambers

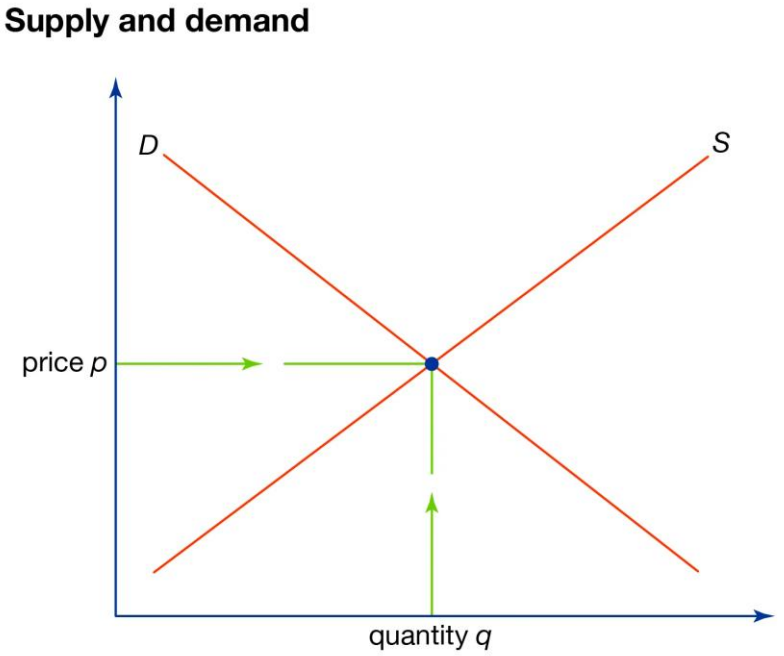
“Common Sense”

- Science has shown repeatedly that the reality of our physical and social worlds is often not what we expect. “Common sense” models often simply don’t reflect reality
- Examples:
  - Echo chambers
  - Minimum wage and employment



“Common Sense”

- Science has shown repeatedly that the reality of our physical and social worlds is often not what we expect. “Common sense” models often simply don’t reflect reality
- Examples:
  - Echo chambers
  - Minimum wage and employment
  - Immigration and wages



“Common Sense”

- Science has shown repeatedly that the reality of our physical and social worlds is often not what we expect. “Common sense” models often simply don’t reflect reality
- Examples:
  - Echo chambers
  - Minimum wage and employment
  - Immigration and wages
  - Shy Trump voters

Modeling

Modeling

- **Data Generating Process (DGP):** Real world events that generate data.
- All data has a data generating process.
- Many of these processes can be represented mathematically.
- These mathematical representations are models. They are not the only kind of model, but they are the kind we will primarily be working with in this class.
- Different kinds of data have different DGPs, and thus merit different models.

Modeling

- **Data Generating Process (DGP):** Real world events that generate data.
- All data has a data generating process.
- Many of these processes can be represented mathematically.
- These mathematical representations are models. They are not the only kind of model, but they are the kind we will primarily be working with in this class.
- Different kinds of data have different DGPs, and thus merit different models.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Sed at turpis vitae velit euismod aliquet. Pellentesque et arcu. Nullam venenatis gravida orci. Pellentesque et arcu. Nam pharetra. Vestibulum viverra varius enim.

Nam laoreet dui sed magna. Nunc in turpis ac lacus eleifend sagittis. Pellentesque ac turpis. Aliquam justo lectus, iaculis a, auctor sed, congue in, nisl. Aenean luctus vulputate turpis. Mauris urna sem, suscipit vitae, dignissim id, ultrices sed, nunc.

Phasellus nisi metus, tempus sit amet, ultrices ac, porta nec, felis. Quisque malesuada nulla sed pede volutpat pulvinar. Sed non ipsum. Mauris et dolor. Pellentesque suscipit accumsan massa. In consectetuer, lorem eu lobortis egestas, velit odio

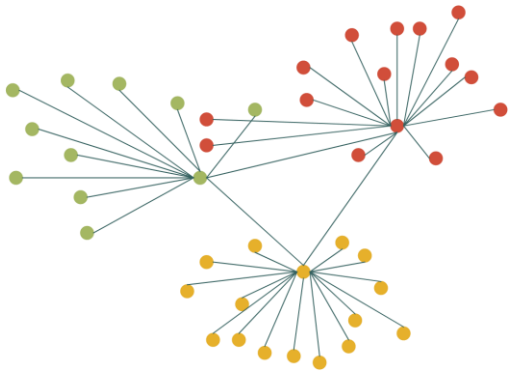
Modeling

- **Data Generating Process (DGP):** Real world events that generate data.
- All data has a data generating process.
- Many of these processes can be represented mathematically.
- These mathematical representations are models. They are not the only kind of model, but they are the kind we will primarily be working with in this class.
- Different kinds of data have different DGPs, and thus merit different models.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Sed at turpis vitae velit euismod aliquet. Pellentesque et arcu. Nullam venenatis gravida orci. Pellentesque et arcu. Nam pharetra. Vestibulum viverra varius enim.

Nam laoreet dui sed magna. Nunc in turpis ac lacus eleifend sagittis. Pellentesque ac turpis. Aliquam justo lectus, iaculis a, auctor sed, congue in, nisl. Aenean luctus vulputate turpis. Mauris urna sem, suscipit vitae, dignissim id, ultrices sed, nunc.

Phasellus nisi metus, tempus sit amet, ultrices ac, porta nec, felis. Quisque malesuada nulla sed pede volutpat pulvinar. Sed non ipsum. Mauris et dolor. Pellentesque suscipit accumsan massa. In consectetuer, lorem eu lobortis egestas, velit odio



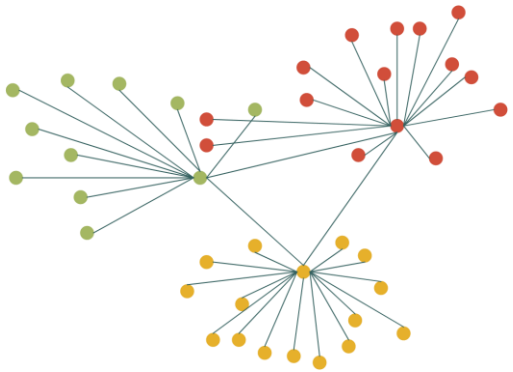
Modeling

- **Data Generating Process (DGP):** Real world events that generate data.
- All data has a data generating process.
- Many of these processes can be represented mathematically.
- These mathematical representations are models. They are not the only kind of model, but they are the kind we will primarily be working with in this class.
- Different kinds of data have different DGPs, and thus merit different models.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Sed at turpis vitae velit euismod aliquet. Pellentesque et arcu. Nullam venenatis gravida orci. Pellentesque et arcu. Nam pharetra. Vestibulum viverra varius enim.

Nam laoreet dui sed magna. Nunc in turpis ac lacus eleifend sagittis. Pellentesque ac turpis. Aliquam justo lectus, iaculis a, auctor sed, congue in, nisl. Aenean luctus vulputate turpis. Mauris urna sem, suscipit vitae, dignissim id, ultrices sed, nunc.

Phasellus nisi metus, tempus sit amet, ultrices ac, porta nec, felis. Quisque malesuada nulla sed pede volutpat pulvinar. Sed non ipsum. Mauris et dolor. Pellentesque suscipit accumsan massa. In consectetuer, lorem eu lobortis egestas, velit odio



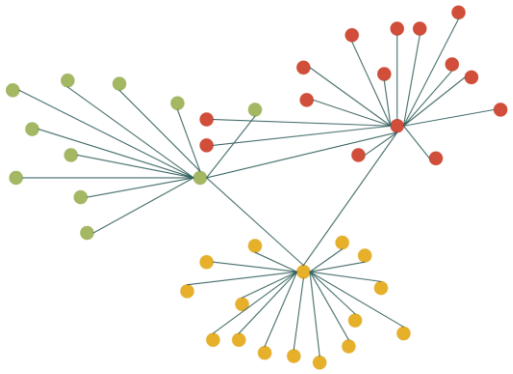
Modeling

- **Data Generating Process (DGP):** Real world events that generate data.
- All data has a data generating process.
- Many of these processes can be represented mathematically.
- These mathematical representations are models. They are not the only kind of model, but they are the kind we will primarily be working with in this class.
- Different kinds of data have different DGPs, and thus merit different models.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Sed at turpis vitae velit euismod aliquet. Pellentesque et arcu. Nullam venenatis gravida orci. Pellentesque et arcu. Nam pharetra. Vestibulum viverra varius enim.

Nam laoreet dui sed magna. Nunc in turpis ac lacus eleifend sagittis. Pellentesque ac turpis. Aliquam justo lectus, iaculis a, auctor sed, congue in, nisl. Aenean luctus vulputate turpis. Mauris urna sem, suscipit vitae, dignissim id, ultrices sed, nunc.

Phasellus nisi metus, tempus sit amet, ultrices ac, porta nec, felis. Quisque malesuada nulla sed pede volutpat pulvinar. Sed non ipsum. Mauris et dolor. Pellentesque suscipit accumsan massa. In consectetuer, lorem eu lobortis egestas, velit odio

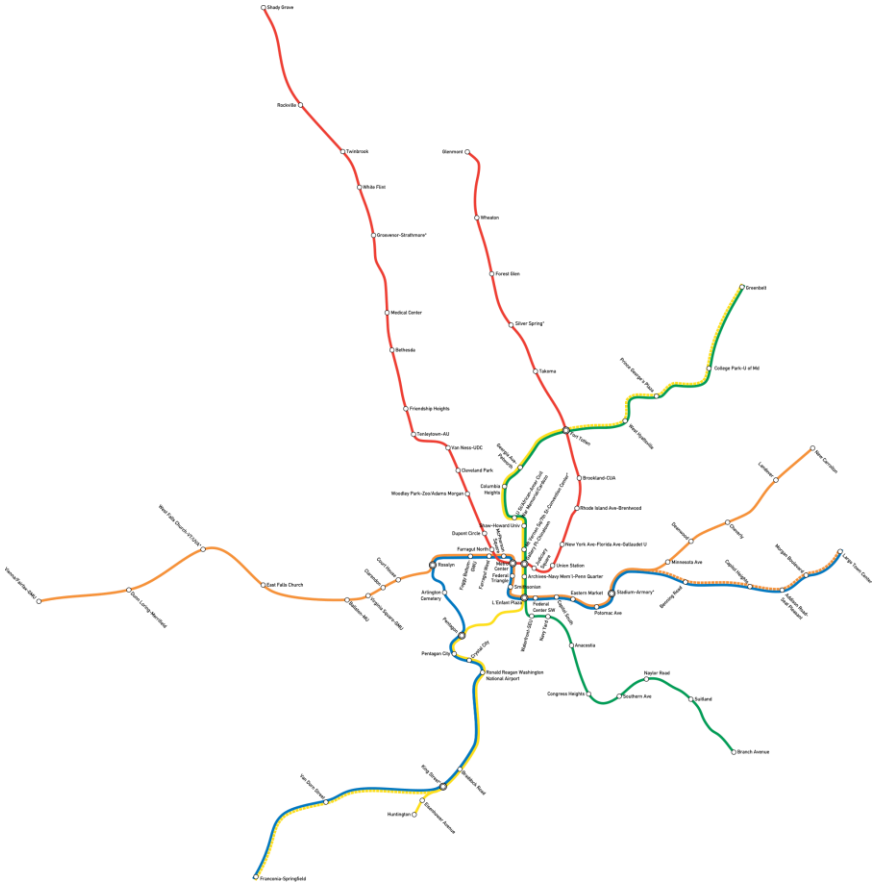


Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

Limitations

A simple model:

Two models of the same system:



Models are useful, but limited

- Endogeneity
- Measurement error
- Spurious correlation
- Misspecification
- Omitted variables
- Simultaneity

Marketing

Why were some of the bad models we discussed successful?

Why were some of the bad models we discussed successful?

- Simple and intuitive

Why were some of the bad models we discussed successful?

- Simple and intuitive
- Appeal to what we want

Attractive ideas are effective ideas

- Writing (and thinking)
- Visualization

Vocabulary

- Data Generating Process
- Model