

Analyzing data

Thomas LaToza



05-899D: Human Aspects of Software
Development (HASD)

Spring, 2011



Today's lecture

Last time

Why would you do a study?

Which type of study should you pick?

What research questions will your study answer?

This time

Should you believe a study's results?

How do you analyze texts?

How do you analyze quantitative data?

Anatomy of a study

Unit(s) of analysis - what you're studying

You have lots of messy data. What thing in data are you studying?

row of data, member of population

bugs, activities using UML class diagrams, design patterns...

Data - what you collected in study

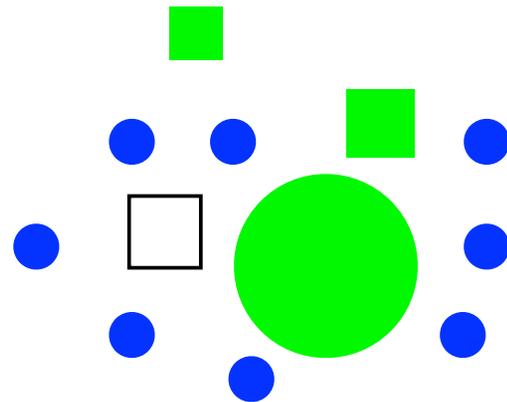
Population - all members that exist

Construct - some property about member

Measure - approximation of construct computed from data

Anatomy of a study of shapes

Real world

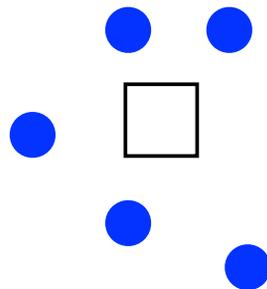


Population

shape
size
filled / empty
color

Constructs

Study



**Sample
of population**

is blue?
size > 10 or size < 10

Measure

(Some) types of validity

Validity = should you believe a result

Construct validity

Does measure correspond to construct or something else?

External validity

Do results generalize from participants to population?

Internal validity (experiments only)

Are the differences between conditions caused only by experimental manipulation and not other variables? (confounds)

Example: Typed vs. untyped languages

S. Hanenberg. (2009). What is the impact of static type systems on programming time? In the *PLATEAU workshop, OOPSLA 09*.

Participants 26 undergrads **Task** write a parser 27 hrs

Setup new OO language 16 hr instructions

Conditions type system vs. no type system
found errors at compile time errors detected at runtime

RESULTS

Developers with untyped version significantly faster completing task to same quality level (unit tests).

Exercise: Validity of type system study

Construct validity

Does measure correspond to construct or something else?

External validity

Do results generalize from participants to population?

Internal validity (experiments only)

Are the differences between conditions caused only by experimental manipulation and not other variables? (confounds)

Other reasons you're skeptical about results?

Good (not perfect) study designs

Goals

Maximize **validity** - often requires more
more participants, data collected, measures
longer tasks
more realistic conditions

Minimize **cost** - often requires
fewer participants, data collected, measures
shorter tasks
less realistic, easier to replicate conditions

Studies are **not proofs** - results could always be invalid

Don't sample all developers / tasks / situations; measures imperfect

Goal is to find results that are

interesting

relevant to research questions

valid enough your target audience believes them

Techniques for qualitative data analysis

Contextual design

Set of models for understanding how work is done

Content analysis / grounded theory

Technique for analyzing texts

Used both to find patterns in data & convert to quantitative data

Process models

Models of steps users do in a task

Taxonomies

What things exist, how are they different, and how are they related?

Affinity diagrams

Technique for synthesizing many disparate observations or interpretations into a coherent whole

Contextual design [Beyer & Holtzblatt]

Analysis technique designed for analyzing contextual inquiry data

But also useful as inspiration for similar models

Types of models

- Flow model - how work is broken up among people & coordinated

 - What information flows between people

- Artifact model - things people create & interact with

 - How structure of artifact influences work

- Cultural model - people & groups that influence how work is done

- Physical model - constraints of physical space

 - Ways in which physical space influences work done

- Sequence model - structure of a task

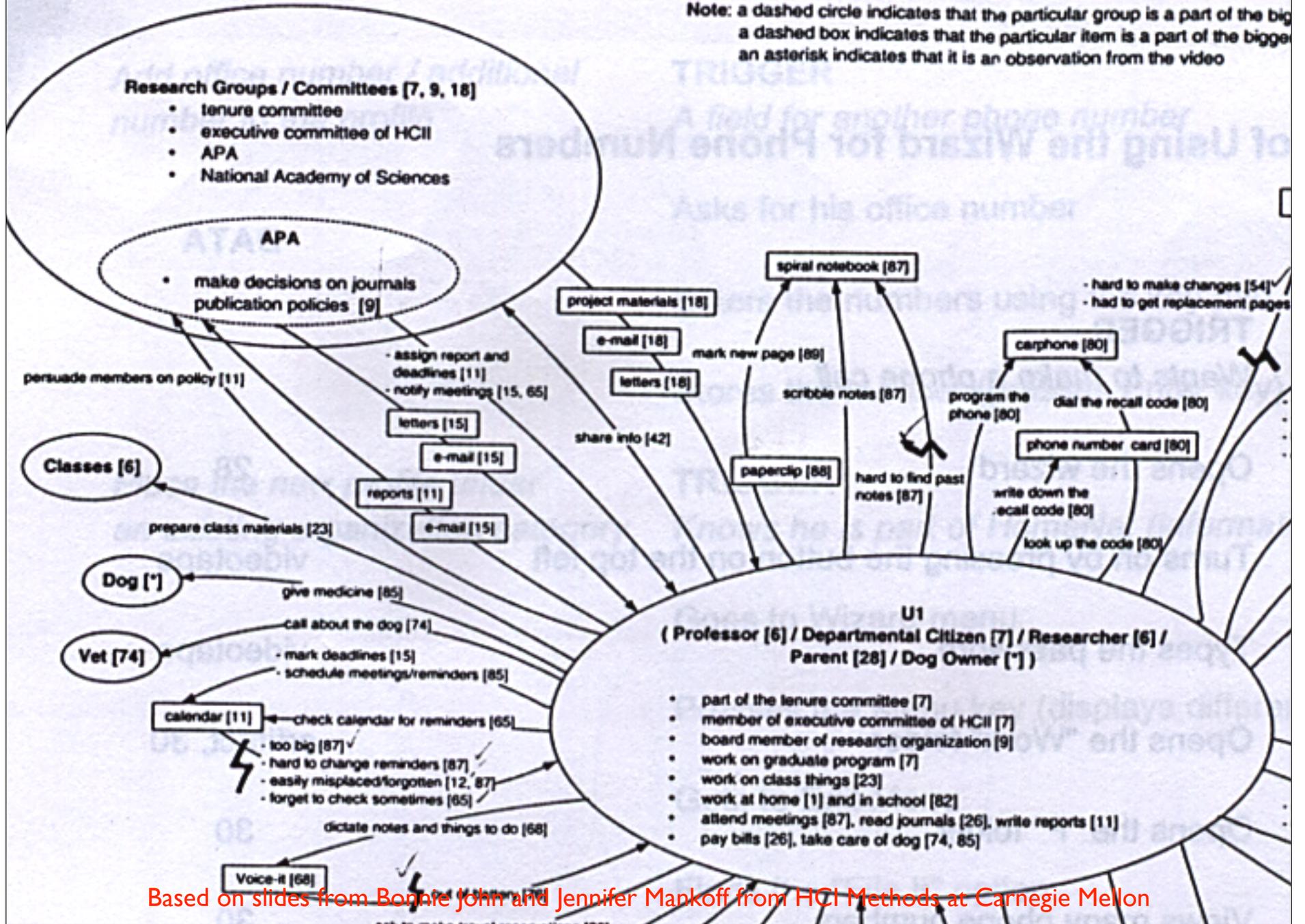
Principles

- Model & interpret work being done

Based on slides from Bonnie John and Jennifer Mankoff from HCI Methods at Carnegie Mellon

Part 1A: Flow Diagram of the User

Note: a dashed circle indicates that the particular group is a part of the big
 a dashed box indicates that the particular item is a part of the bigger
 an asterisk indicates that it is an observation from the video



Based on slides from Bonnie John and Jennifer Mankoff from HCI Methods at Carnegie Mellon

Can see whole month at once [85]

Only special meetings are marked,
not everyday occurrences like classes [85]

Some entries are circled, some entries are not [*]

Opened to the current month [*] ✓ *GREAT observation!*

Special lunch meetings, class meetings, visitors are marked [85] ✓

Special reminders: heart on 1st of month to remind her to give
medicine to dog [85] ✓

Big enough spaces to make notes [85]

Lots of notations in date cells [85]

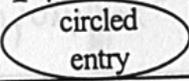
Trips are also marked [85]

↳ Date cells can be too small to fit all reminders [87]

↳ Tend to leave it (forget it) [87]

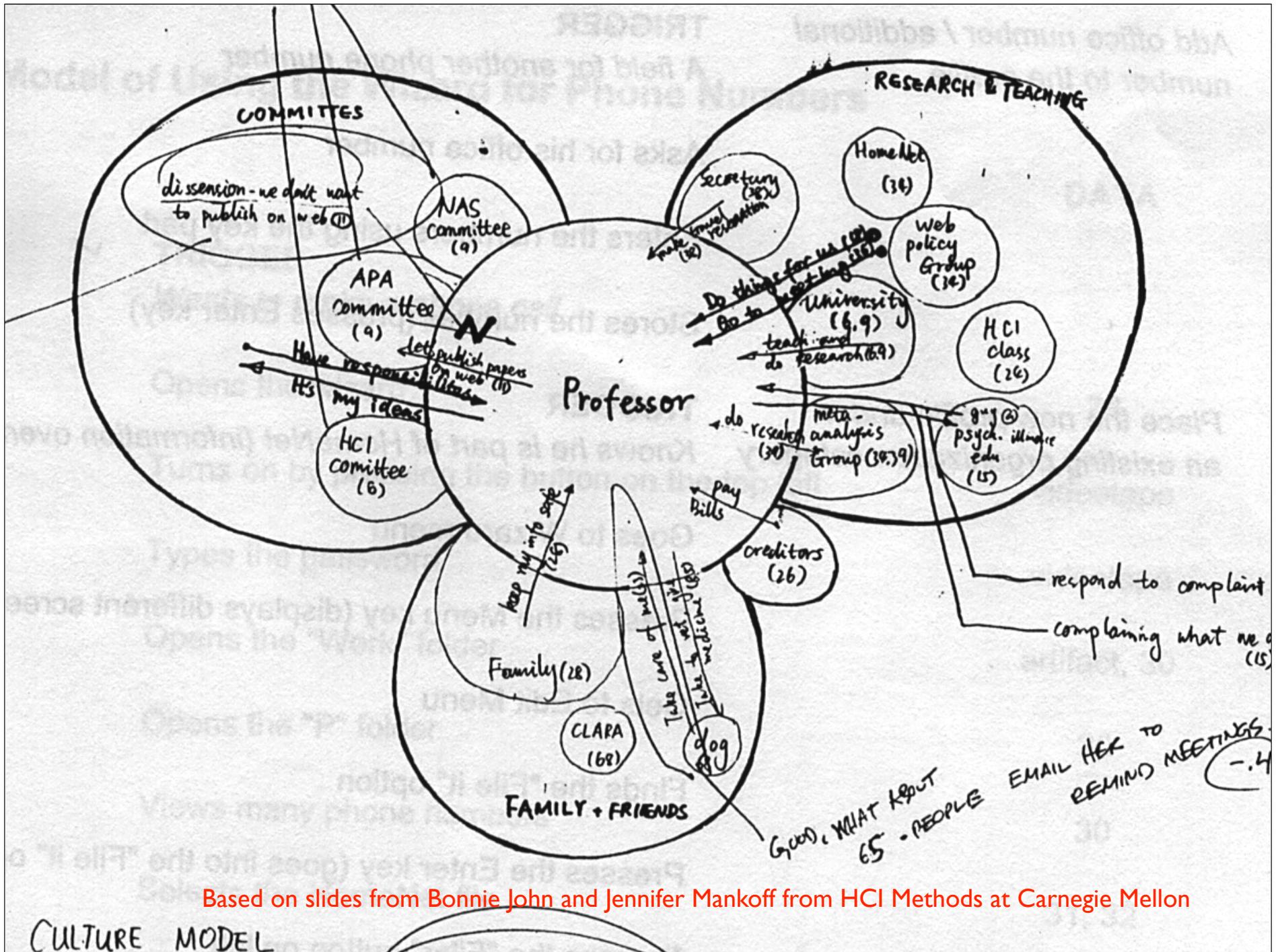
↳ Can't change or rearrange entries; has to scribble things out [87] ✓

↳ Too big (won't fit in pocket/purse) [87] ✓

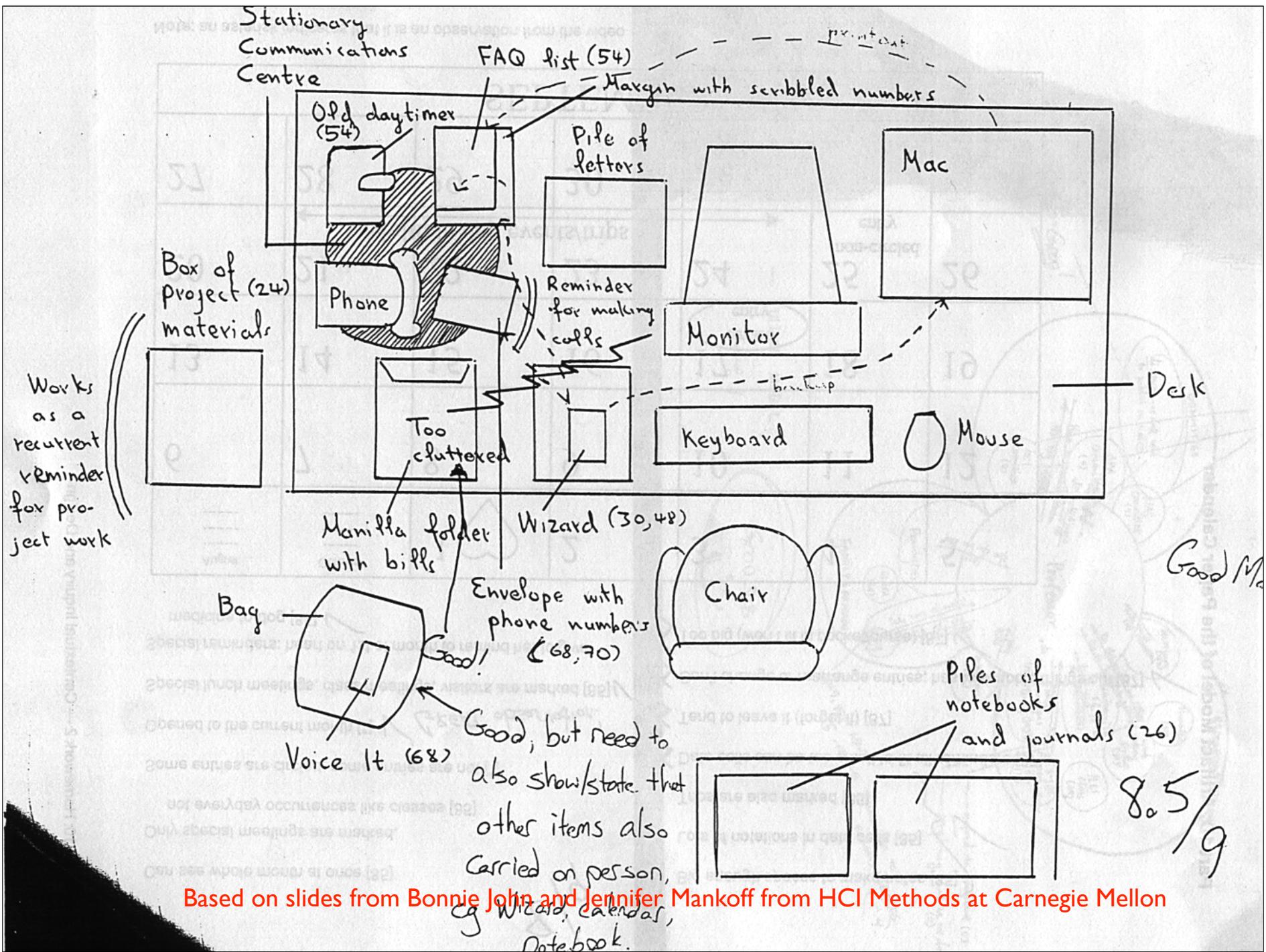
August ----- ----- -----	October ----- ----- -----	1 	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17 	18	19
20	21	22	23	24	25 non-circled entry	26
27	28	29	30	← Multiple day events/trips →		
SEPTEMBER						

Based on slides from Bonnie John and Jennifer Mankoff from HCI Methods at Carnegie Mellon

that it is an observation from the video



Based on slides from Bonnie John and Jennifer Mankoff from HCI Methods at Carnegie Mellon



Based on slides from Bonnie John and Jennifer Mankoff from HCI Methods at Carnegie Mellon

Intent - what is sequence trying to achieve

Trigger - thing that causes sequence to begin

Step - action or thought preceding action

Arrows - order of steps (may have loops)

Lightning bolt - breakdown that occurred

INTENT	TRIGGER	DATA
<i>Find a phone number</i>	TRIGGER <i>Wants to make a phone call</i>	
	Opens the wizard	28
	Turns on by pressing the button on the top left	videotape
	Types the password	videotape
	Opens the "Work" folder	artifact, 30
	Opens the "P" folder	30
	Views many phone numbers	30
	Selects the HomeNet file	31, 32
	Selects a phone listing	32
	States the file may be old (⚡ doesn't update info)	32
<i>Check to see if the phone number is in the Wizard</i>	TRIGGER <i>Get an e-mail with phone number in it</i>	43
	Visually looks at the Wizard	44
	Opens the "Work" phone listing	44
	Finds the phone number corresponding to this name	44
	Asks if it is correct	44
<i>Change the phone number to the correct one</i>	TRIGGER <i>The phone number in the wizard is incorrect</i>	45
	Presses the Enter key to open a listing	46
	Interviewer's name and phone number displayed in a single field	46
	Get the new phone number	47
	Types the new number using the key pad	48

Based on slides from Bonnie John and Jennifer Mankoff from HCI Methods at Carnegie Mellon

Quantitative analysis of text

Examples of textual data:

think-aloud transcripts, emails, bug descriptions, source code

Goal: quantitative analysis

		[??], hm that's an interesting little side effect		
		Uhm,		
Scrolling	BufferHandler.delayUpdate()+	wonder how much time I've got, let's see. [14 15 mintues in]		
Scrolling				
Scrolling	BufferHandler.delayUpdate()+	I'm not allowed to finish this after 15 minutes.		
ReferencesTo	BufferHandler.delayUpdate()+			
		Ok. Well code duplication is evil. [??]		
		There is clearly a reason you want to delay the firing of the event		
		But you eventually need to fire the event so that someone else receives that		
		And that seems reasonable		
		So, it's absurd to have a getter that's a mutator		
		It's absurd to have a getter that doesn't use it's return value		
		Which means I need a separate method which has a void return		
		which does the mutation and then I need to reconcile that with the		
		getter so I don't have code duplication		
		And it's architecturally conservative, it's unlikely to introduce a bug		
		Yeah, let's see is there any other substance to this complaint		
		getFoldLevel() is both architecturally questionable and clearly bad design		
		architecturally it is intended to change the buffers state from within a different		
		component.		
		You see I would argue that trivially speaking this is a state change		
		but it's purpose in life is really to fire an event, conservatively?		
		[writing]		

Content analysis

1. **open** coding - read through the text
 - look for **interesting** things relevant to research questions
 - add notes in the margin (or column of spreadsheet)
 - add “**codes**” naming what you saw
 - make up codes as you go, not systematic
2. **axial** coding - how are codes related to each other?
 - look for **patterns**: causality, ordering, alternatives
3. **selective** coding - from initial codes, select interesting ones
 - which codes found interesting things?
 - from initial examples, build definition on when they are applied
 - systematically** reanalyze data and apply codes
4. **second** coder (optional)
 - 2nd person independently applies codes from definitions
 - check for interrater **reliability** - if low, iterate defns & try again

Example [LaToza+ FSE07]

Research questions

What is **process** developers go through in making a code change?
 How does developer **experience** influence this process?

		[??], hm that's an interesting little side effect					
Scrolling	BufferHandler.delayUpdate()+	Uhm,					
Scrolling		wonder how much time I've got, let's see. [14 15 mintues in]					
Scrolling	BufferHandler.delayUpdate()+	I'm not allowed to finish this after 15 minutes.					
ReferencesTo	BufferHandler.delayUpdate()+	[redundant]					
Explain		Ok. Well code duplication is evil.	[??]	He already seems to know that you don't want to fire the event frequently			
		There is clearly a reason you want to delay the firing of the event		WHY???			
		But you eventually need to fire the event so that someone else receives that					
		And that seems reasonable					
Critique		So, it's absurd to have a getter that's a mutator					
Critique		It's absurd to have a getter that doesn't use it's return value					
DPC	Add new force update method	Which means I need a separate method which has a void return					
		which does the mutation and then I need to reconcile that with the					
		getter so I don't have code duplication					
		And it's architecturally conservative, it's unlikely to introduce a bug					
		Yeah, let's see is there any other substance to this complaint					
		getFoldLevel() is both architecturally questionable and clearly bad design					
		architecturally it is intended to change the buffers state from within a different				Should have said private state...	
		component.		Should have said breaks info hiding...			
Rationale	why does doDelayedUpdate call getFoldLevel	You see I would argue that trivially speaking this is a state change					
Explain		but it's purpose in life is really to fire an event, conservatively?			!!!!		
		[writing]					

Process models

How do users go about a task?

Intent: what is their goal?

Step: what action did they do? how much time? does it succeed?

Breakdowns: problems that occur - wasted work, unable to do something

Intent important to capture - less affected by alternative approaches

Example [Stylos & Clarke, ICSE07]

Two conditions:

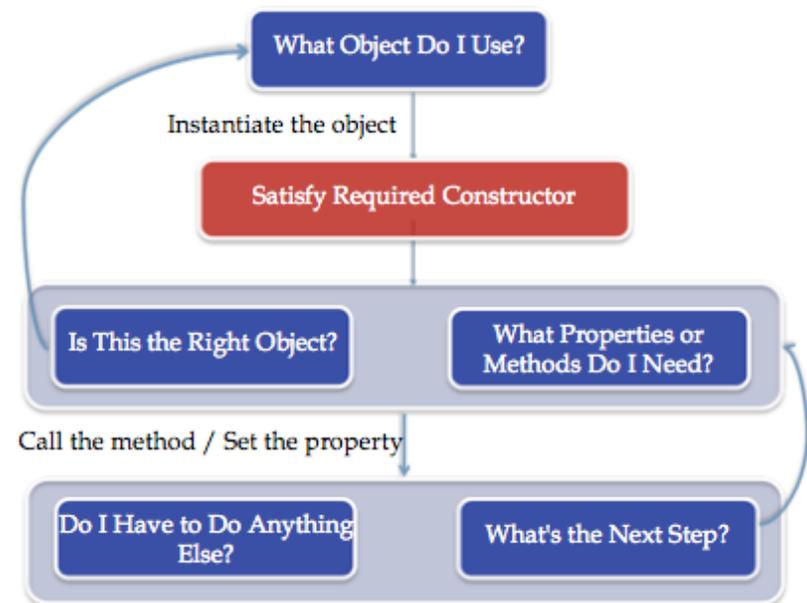
```
var foo = new FooClass();  
foo.Bar = barValue;  
foo.Use();
```

Default constructor ("create-set-call")

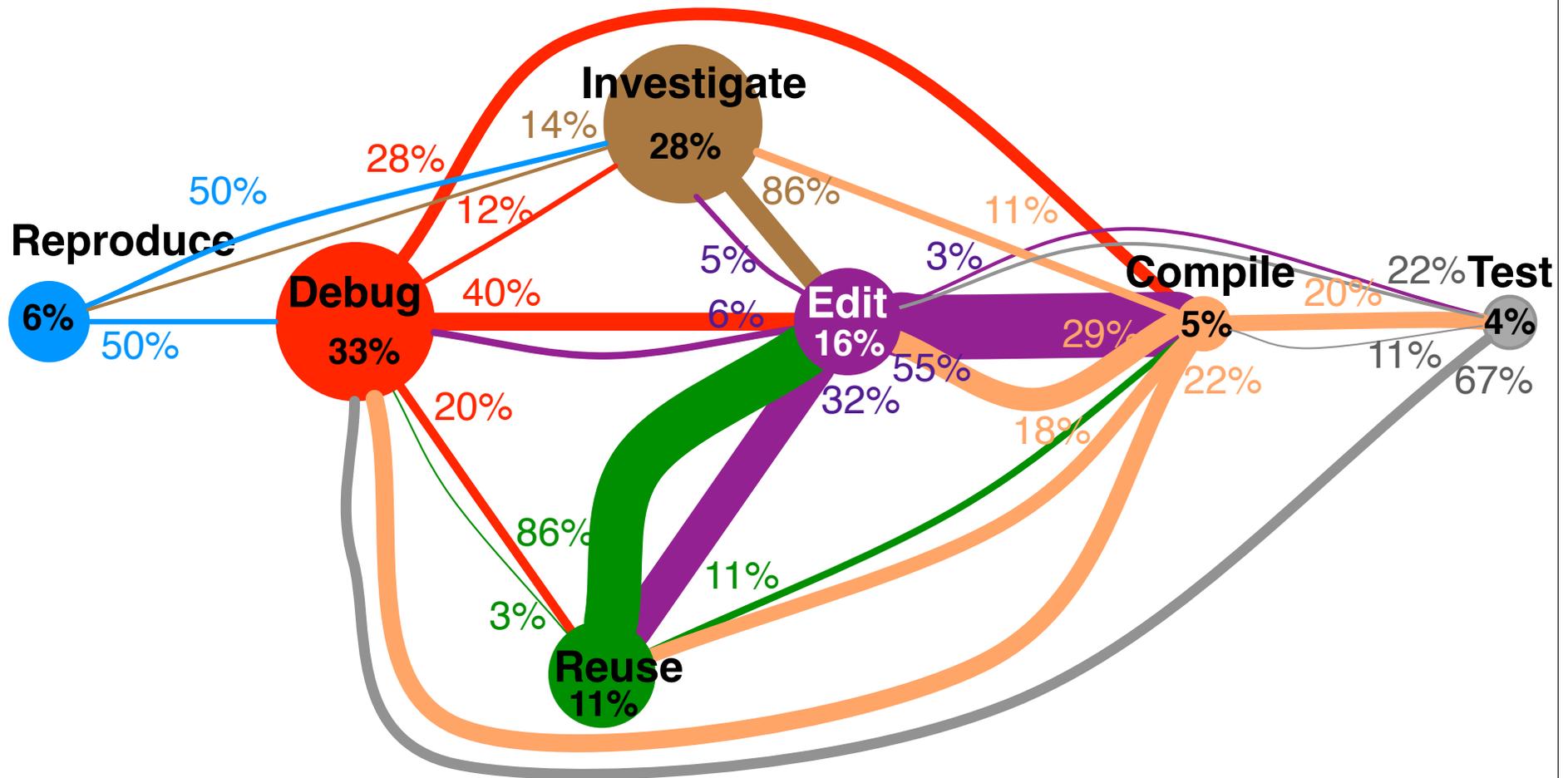
vs

```
var foo = new FooClass(barValue);  
foo.Use();
```

Required constructor



Example 2 - Developer coding activity



Circle size % of time **Edge thickness** % of transitions observed

Thomas D. LaToza and Brad A. Myers. 2010. [Developers ask reachability questions](#). In Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume 1(ICSE '10), Vol. 1. ACM, New York, NY, USA, 185-194.

Taxonomies

What types of things **exist**? (e.g., biologists identify species)

Build definitions - what qualities or properties do these things have?

Distinguish definitions - what property(s) differ?

Relationships often qualitative relationships based on how related

Model usually associated with content analysis

Example [LaToza+ FSE07] - types of **facts** in code

Changeable - developers thought **alternatives** existed

Constraints - **changing** fact would change other facts

Frozen - cannot change fact

Hypotheses (speculation) or observed (seen in code)

Hacks - **bad** design, should be changed

True or false

Code (about implementation) or requirements

Overlooked (**unintentional** by original dev) vs. intended

Affinity diagram

How do you consolidate many observations / interpretations?

Bottom up analysis of **data**

- Interesting snippets in transcript

- Experimenter notes from sessions

- Interpretation attached to a piece of data

- Notes independently generated from multiple experimenters

- Goal: LOTS (100s, 1000s) of small pieces of data

Process

- Write / Print data snippets onto 3 x 5 index cards

- Place one by one on a large wall

- Place cards next to similar cards to form topic clusters

- Label clusters with topic sticky notes

- Rearrange cards as new topics emerge

- Use hierarchical organization of topics to structure discussion of data

Example [LaToza+ ICSE06]



- so fundamental everyone has modified
Some files ~~are~~ almost only person checks in
others where everybody checks in

VS

Started out w/ buddy ownership
(co-owning)

- also for code reviews

broke down close to ship

XML - S5 - Gina

Dev owns code for feature
work

BIZ - S10 - Gina

Not OK to work in someone
else's code

MAP - S2 - Gina

"Any time you own something
you own it forevermore"

VS

Code ownership based on
features

- some overlap
- assignit based on skills,
interests

into ownership XML

ty on owned code,
of knowledge, only
their component
w/ code, less ownership

XML - S5 - Gina

cross into other's
quick bug fixes

code ownership

g fixing

ancing

area of code VS

e
another product unit

out of area to fix

Qualitative vs. quantitative

Qualitative analysis most useful for

Figuring out what's there, what's being done

What's important?

How are things done?

Why is person doing / using / thinking something?

Limitations

Small n: few examples, may not generalize

Interpretations could be biased

Quantitative analysis most useful for

Testing hypotheses

Investigating relationships between variables

Predicting

Limitations

Lack interpretation

Therefore, great quantitative studies also do qualitative analysis

Types of quantitative data analysis

Frequency

How often do things occur? (counts, %s, avg times, ...)

Correlational

How are multiple variables related? (correlations, ...)

Estimate variable x from variables a, b, c (regression, classifiers, ...)

Controlled experiment (causal)

Does a change in x **cause** a change in y ?

Studies often use more than one

Sampling a distribution

You have a sample of a population

How likely are results on sample to differ from true population?

Solution: **statistics!**

True population is described by a distribution (gaussian, poisson,)

You took an **unbiased** sample of distribution to generate sample

Infer parameters of distribution from your sample

Use statistical formulas from distribution to

- accept or reject hypothesis

- assign a range of possible values real population might have

- understand how variables are related

Frequency analysis

Good for

- % of time, questions, actions, activity, ...

- Range of times to do something

Range - minimum & maximum value

- Report confidence interval or standard deviation for uncertainty

- E.g., 95% of time, true value will fall in this range

- Can report numbers or plot w/ box plots

Measures of central tendency

- Mean: meaningful for distributions that aren't skewed

- Mode: most frequent

- Median: midpoint value - better for skewed distributions

Useful to plot distribution of your data to see what it looks like

Correlations

How is variable x related to variable y?

-1 variables are perfectly inversely related

0 no relation at all

1 variables are always the same

Pearson's r:

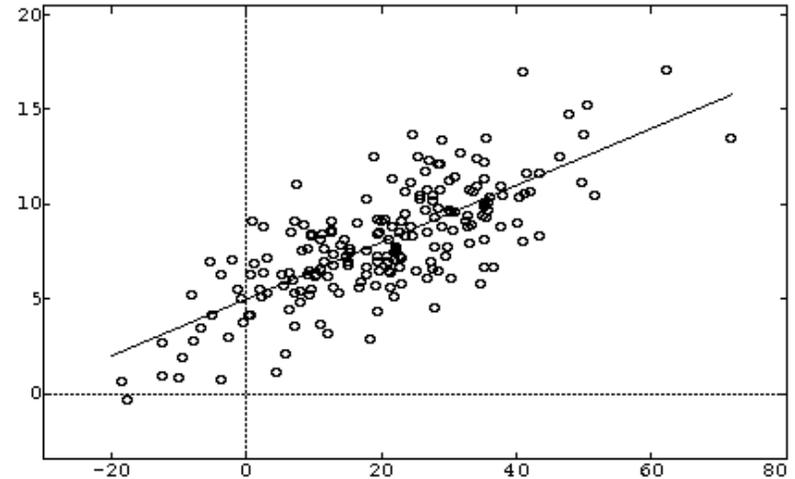
$$r_{xy} = \frac{\text{SUM}(x_i - x_{\text{mean}})(y_i - y_{\text{mean}})}{(n - 1) \text{stdev}_x * \text{stdev}_y}$$

Has assumptions about nature of data in order to valid

Significance: how likely is this relationship do to chance?

Regression

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i,$$



Predict variable y using variables x_0, \dots, x_n

Useful for predicting y from other variables in data

E.g., predict number of defects in code from features of code...

Infer best coefficients β_0, \dots, β_n

Makes assumptions about nature of data, but fancier forms w/ fewer assumptions

Controlled experiment

Only way to argue **causality** - change in var x causes change in var y

Manipulate **independent** variables

Creates “conditions” that are being compared

Can have > 1 , but # conditions usually exponential in # ind. variables

Measure dependent variables (a.k.a measures)

Quantitative variable you calculate from collected data

E.g., time, # questions, # steps, ...

Randomly assign participants to condition

Ensure that participants only differ in condition

Not different in other **confounding** variables

Test hypotheses

Change in independent variable causes dependent variable change

E.g., t-test, ANOVA, other statistical techniques

Between subjects or within subjects

Between subjects - participants do **one** condition

Compare between participants in different conditions

Better for

- small participant differences

- large ind variable effects

- longer tasks

Within subjects - participants do **multiple** conditions

Compare between same participants in different conditions

Better for

- large participant differences

- small ind variable effects

- small interactions between tasks

 - (task 1 doesn't make task 2 easier)

Use counterbalancing - 1/2 get task 1 first, 1/2 get task 2 first

prevents confound of order

Exercise: design a study to....

Test if a new static analysis tool that finds bugs helps Java developers to be more productive than using Eclipse

Resources

Which statistical test should I use?

<http://www.muhlenberg.edu/depts/psychology/docs/StatsDecision.pdf>

Recommended statistical package

JMP - www.jmp.com - designed for exploratory data analysis

General reference for quantitative data analysis

Robert Rosenthal & Ralph Rosnow. (2007). Essentials of Behavioral Research: Methods and Data Analysis. McGraw-Hill.

General reference for qualitative data analysis

Michael Quinn Patton. (2002). Qualitative Research & Evaluation Methods. Sage Publications.

Questions?