Using Computational Methods to Analyze Educational Data

Camilo Vieira, PhD
Alejandra J. Magana , PhD
Mireille (Mimi) Boutin, PhD

49th Annual Frontiers in Education (FIE) Conference
October 19th, 2019

Informática Educativa

This work has been partially supported by the NSF grant awards EEC#1826099.

About Us

Dr. Camilo Vieira

Assistant Professor at Universida del Norte (Colombia)

Computational Science and Engineering Education

Computing education

Learning analytics for engineering education.

Prof. Alejandra J. Magana

Professor of the Computer and Information Technology Department at Purdue

Characterizing modeling and simulation practices in science and engineering

Cyberlearning in STEM

Prof. Mireille “Mimi” Boutin

Associate professor in the School of Electrical and Computer Engineering at Purdue

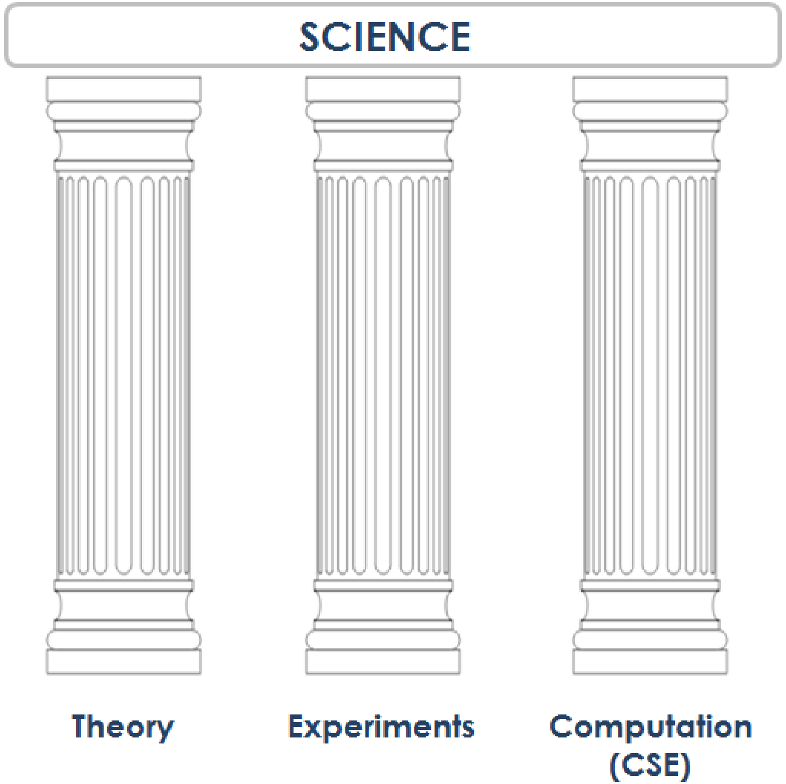
“Learning by Teaching”
www.projectrhea.org

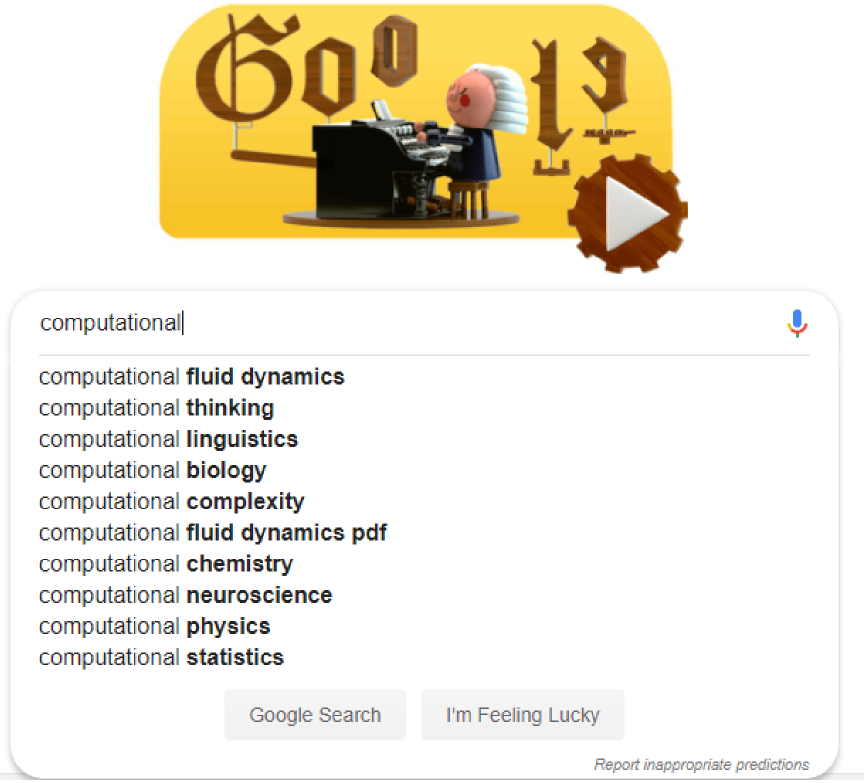
Machine learning and applied mathematics

The plan for today

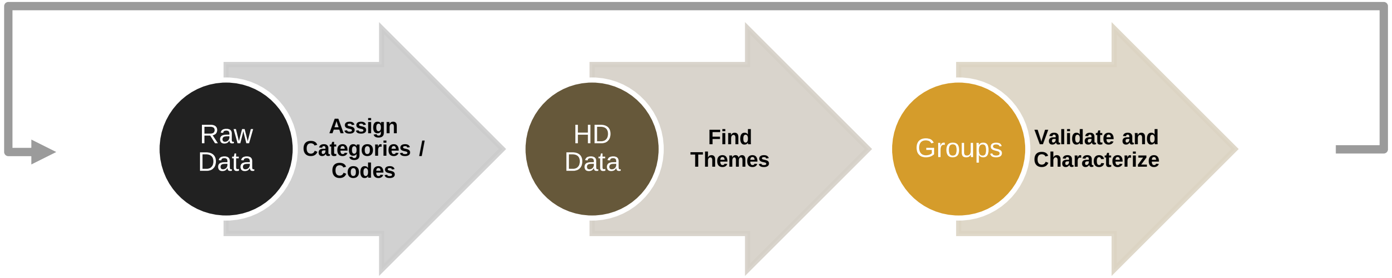
1. Introducing the studies (45 minutes)
2. Conforming groups (5 minutes)
3. R Programming (25 minutes)
4. Wrap up (5 minutes)

“For every professional software developer in the world, there are ***nine more people programming as part of their jobs who aren't professional software developers****” Guzdial, 2015.*


SCIENCE
Theory
Experiments
Computation (CSE)


computational
computational fluid dynamics
computational thinking
computational linguistics
computational biology
computational complexity
computational fluid dynamics pdf
computational chemistry
computational neuroscience
computational physics
computational statistics
Google Search
I'm Feeling Lucky
Report inappropriate predictions

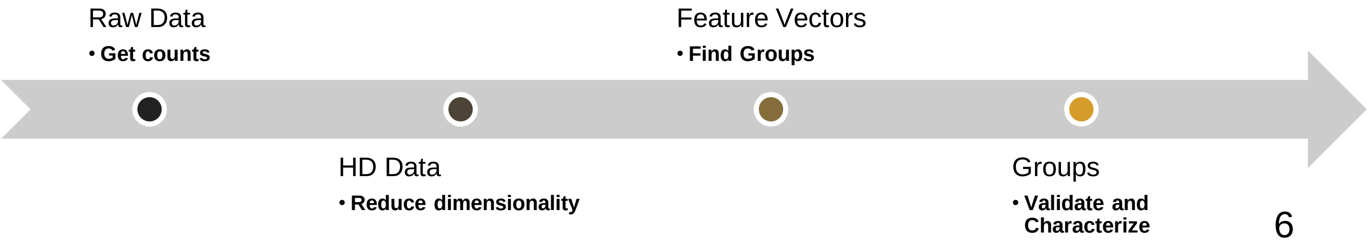
A Data Analysis Process

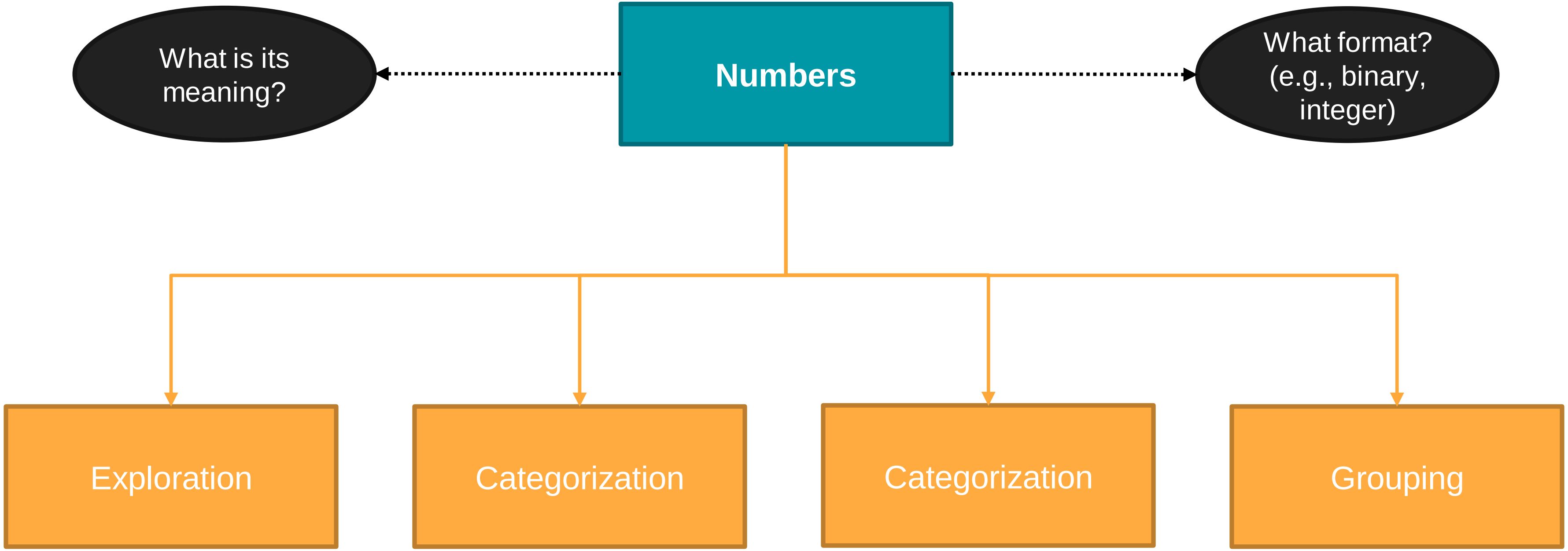


Based on **numbers,** supports

Exploration | Categorization | Grouping | Validation

A Pattern Recognition Process




What is its meaning?
Numbers
What format? (e.g., binary, integer)
Exploration
Categorization
Categorization
Grouping


CAUTION
Computers DON’T do all the work

Exploration | Categorization | Grouping | Validation

Literature Review

Summarize findings of a literature review using visualizations

VISUAL LEARNING ANALYTICS – A Lit. Review

Goal:
(1) Identify the main uses of visualizations in educational research
(2) Identify gaps and future opportunities for conducting research and creating visualization tools that can help advance the field of visual learning analytics.

Hypothesis: InfoVis experts and Education experts working separately.

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, *122*, 119-135.

Literature Review

Traditional Approach:

Table 2. Classification of case studies according to the learning settings

Learning setting[*]	Authors & Year (Paper Ref.)
VLEs / LMSs[a]	Lin, Hsieh & Chuang, 2009; Lykourentzou et al., 2009a; Lykourentzou et al., 2009b; Macfadyen & Dawson, 2010; Merceron & Yacef, 2008; Romero et al., 2008; Romero-Zaldivar et al., 2012; Tanes et al., 2011
MOOC/social learning[b]	Clow & Makriyiannis, 2011; Fournier et al., 2011; Kizilcec et al., 2013
Web-based education[c]	Abdous, He & Yen, 2012; Giesbers et al., 2013; He, 2013; Khribi et al., 2009; Li et al., 2011; Romero et al., 2009
Cognitive tutors[d]	Baker et al., 2008; Moridis & Economides, 2009; Pardos et al., 2013; Shih, Koedinger & Scheines, 2008
Computer-based education[e]	Ali et al., 2012; Barla et al., 2010; Blikstein, 2011; Jeong & Biswas, 2008; Levy & Wilensky, 2011; Santos et al., 2012; Thai-Nghe et al., 2011
Multimodality[f]	Worsley & Blikstein, 2013
Mobility[g]	Chen & Chen, 2009; Leong et al., 2012

VISUAL LEARNING ANALYTICS

Criteria	1-2	…	9-10
Connection with Visualization Background (CVB)	The paper does not mention any visualization theory, guidelines, or principles		The paper critically analyzes the relevant literature in visualization and makes an informed decision for the proposed tool and discusses the findings using existing literature in visualization

VISUAL LEARNING ANALYTICS

Criteria	1-2	…	9-10
Connection with Visualization Background (CVB)	The paper does not mention any visualization theory, guidelines, or principles		The paper critically analyzes the relevant literature in visualization and makes an informed decision for the proposed tool and discusses the findings using existing literature in visualization
Connection with Educational Theory (CET)	The paper does not mention any educational theory		The paper engages in critical analysis of educational theory and informs the design of the visualization using an educational theory and discusses the results of the work under the lens of this theory

VISUAL LEARNING ANALYTICS

Criteria	1-2	…	9-10
Connection with Visualization Background (CVB)	The paper does not mention any visualization theory, guidelines, or principles		The paper critically analyzes the relevant literature in visualization and makes an informed decision for the proposed tool and discusses the findings using existing literature in visualization
Connection with Educational Theory (CET)	The paper does not mention any educational theory		The paper engages in critical analysis of educational theory and informs the design of the visualization using an educational theory and discusses the results of the work under the lens of this theory
Sophistication of Visualization (SoV)	None of these elements is present: (a) Multiple visualizations; (b) Connected visualizations; (c) Visualizing data at multiple levels; (d) Interactive Visualization (e) Novel Visualization		All these elements are present: (a) Multiple visualizations; (b) Connected visualizations; (c) Visualizing data at multiple levels; (d) Interactive Visualization (e) Novel Visualization

https://infovisreview.github.io/VLAVisualizations/

Exploration | Categorization | Grouping | Validation

Classroom Observation

Characterize changes or improvements in teaching practices via classroom observations

Coding for Kids

Goal: Identifying changes in teaching practices from longitudinal data

Preparing Mentors

- 3 Days
- Technical Sessions
- Lesson Plans
- Mentoring Process

Preparing Teachers

- 2-day Regional Workshops
- Day One: Modeling Lesson Plans 1 & 2
- Day Two: Practice Teaching – Lesson Plans 3 & 4

In-place Mentoring x 4

- ½ day school visit
- Classroom Observation
- Feedback and Mentoring

Closing Events

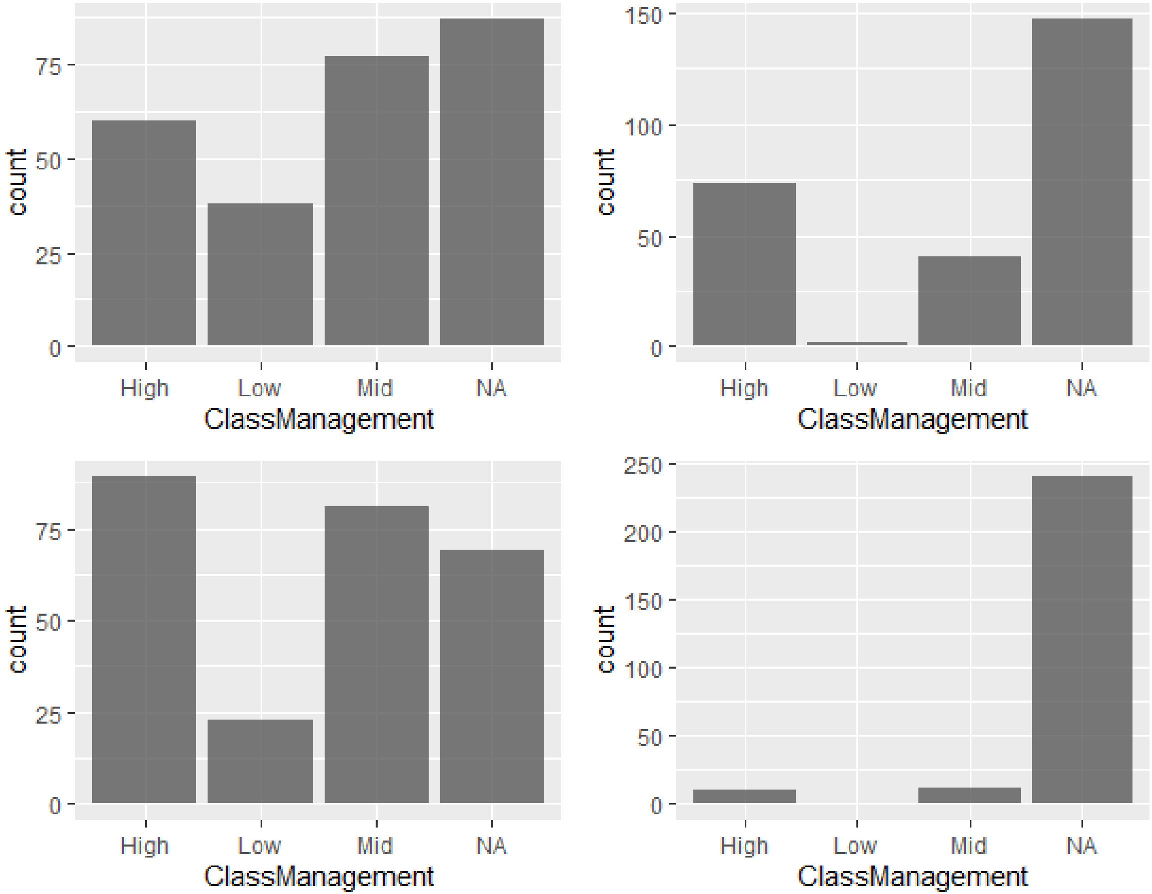
- Closing Workshop for Teachers
- CodeFest for Students

Rubric for Classroom Observation

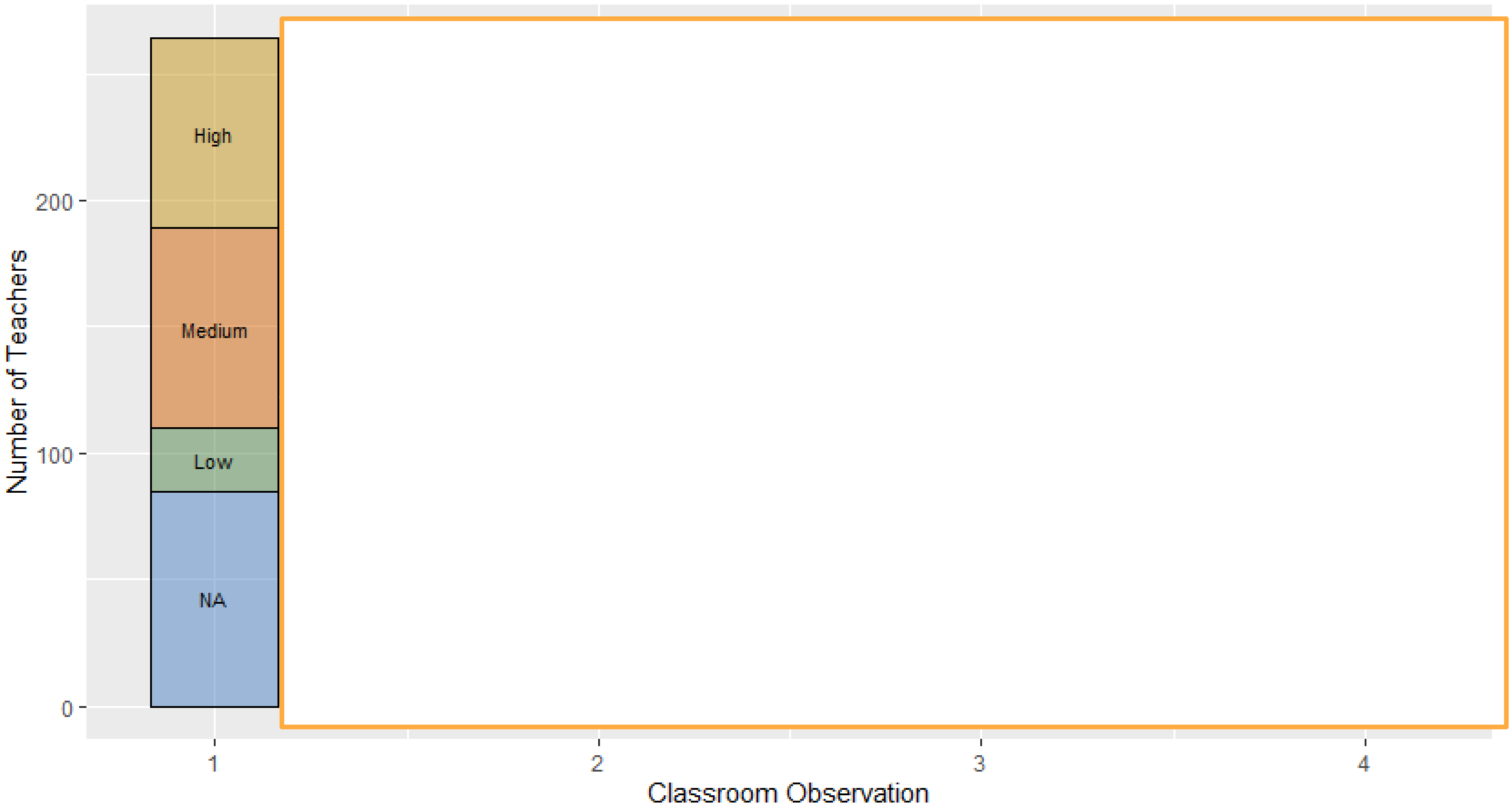
	Low	Medium	High
Norms and Procedures	There are no clear norms and procedures. The instructor has to ask for silence often, and repeats frequently how to get the materials.	There are some norms but students do not always follow them, and the teacher does noting to promote it.	There are clear norms and procedures, which students follow most of the time.

Distribution and Longitudinal Data

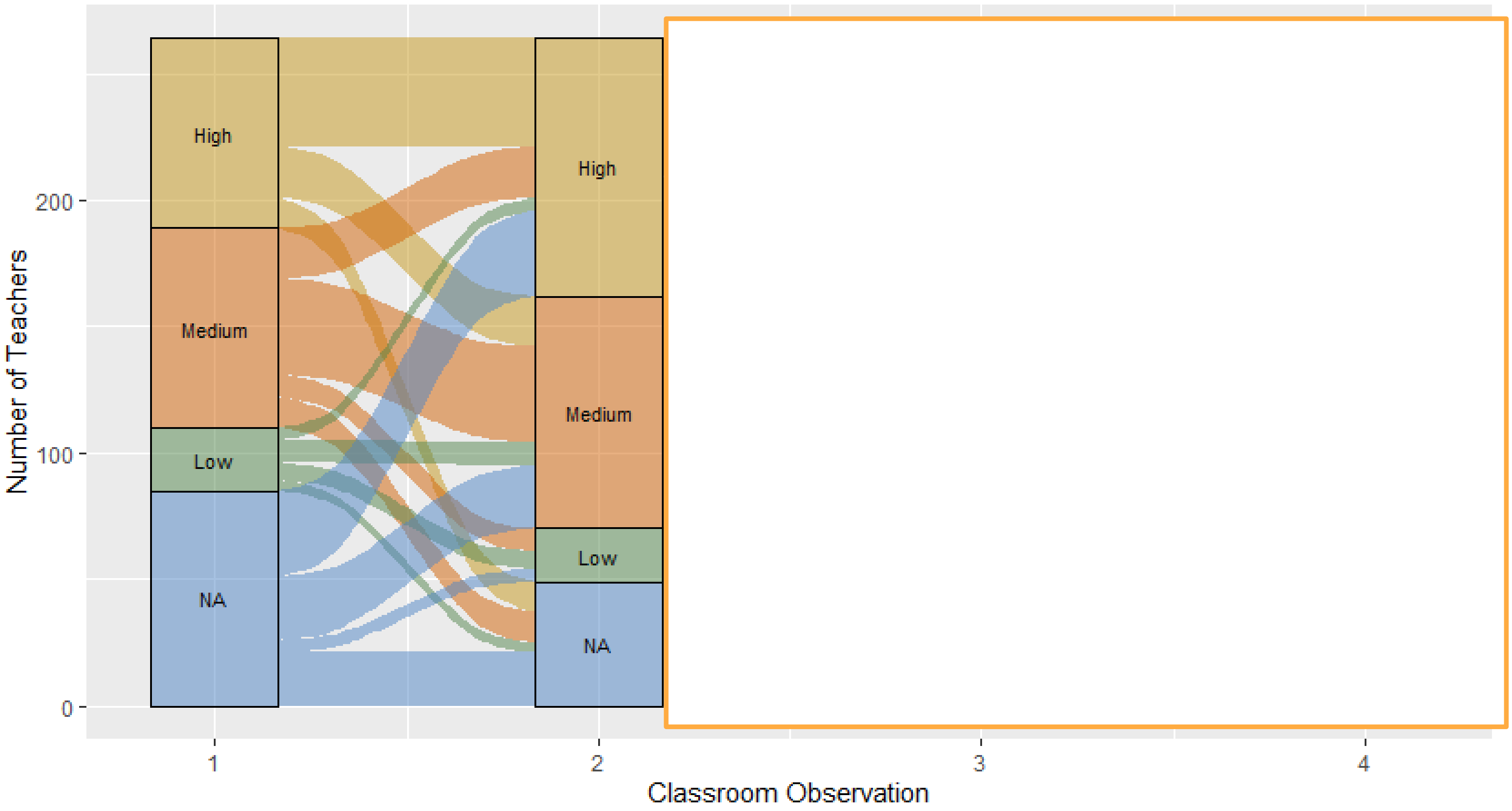
Traditional Approach:



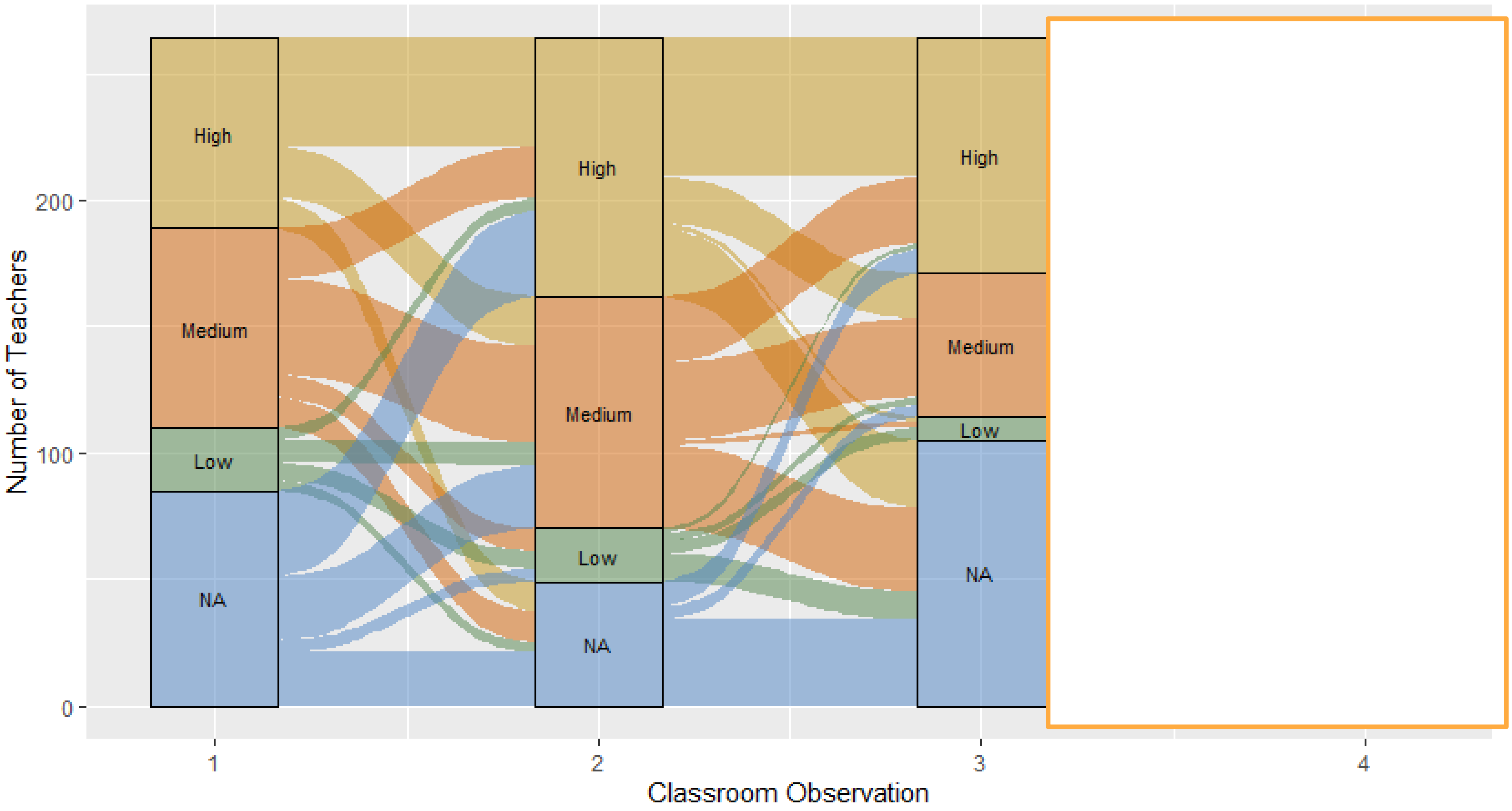
Progression on Effective Teaching Time



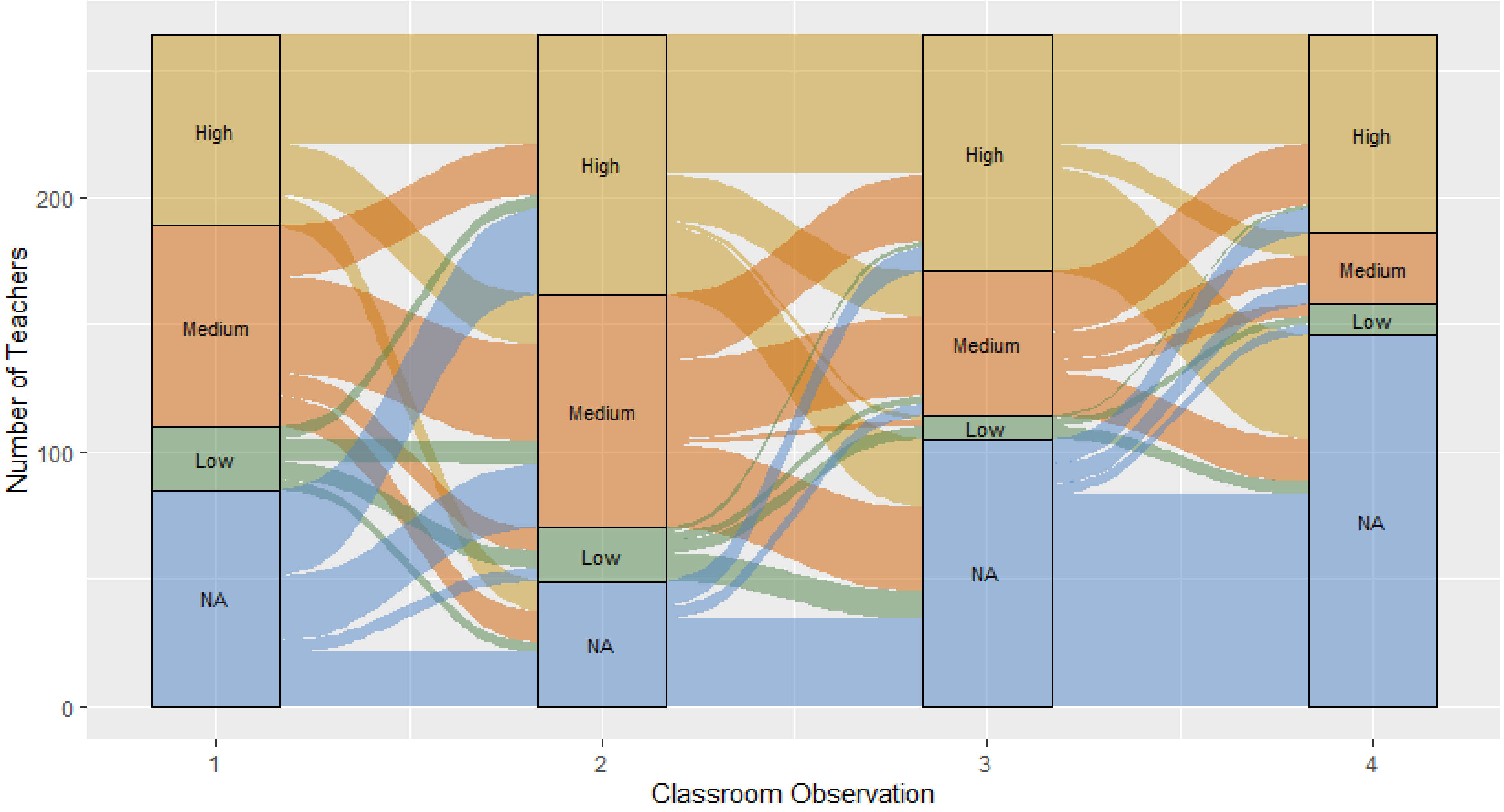
Progression on Effective Teaching Time



Progression on Effective Teaching Time



Progression on Effective Teaching Time



Exploration | Categorization | Grouping | Validation

Think-Aloud Data

Identify patterns of types of knowledge students used in different stages of the problema solving process

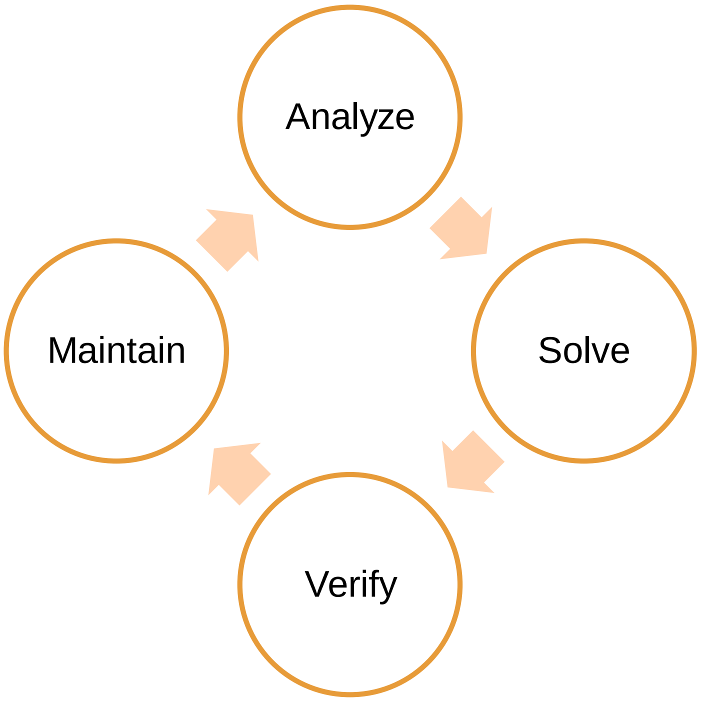
Analysis of think-aloud data

Context:

- **Programming course** for Materials Science and Engineering
- 17 students were invited to participate in a think-aloud activity doing **computational modeling**

Goal: Characterize students' modeling process and usage of the different types of knowledge

An adaptation of the modeling process (Shiflet & Shiflet, 2014).



Types of Knowledge (ToK)

Type of knowledge (Shavelson, Ruiz-Primo, Li, & Ayala, 2003)
Declarative (knowing that)
Procedural (knowing how)
Schematic (knowing why)
Strategic (knowing when, where, how)

Sample counts: **ToK x Modeling step by Student**

Student 1				
	Analyze	**Solve**	**Verify**	**Maintain**
Declarative	10	7	2	0
Procedural	0	13	0	1
Schematic	0	0	1	0
Strategic	1	6	1	3

Sample counts: **ToK x Modeling step by Student**

Traditional Approach:

2014

Student 1 (Milan_L)				
	Analyze	Solve	Verify	Maintain
Declarative	7	0	1	0
Procedural	1	5	1	0
Schematic	1	1	1	0
Strategic	1	4	1	0

Student 2 (Lennon_L)				
	Analyze	Solve	Verify	Maintain
Declarative	9	3	2	1
Procedural	1	8	2	0
Schematic	2	0	0	1
Strategic	4	6	5	2

Student 3 (Charlie_H)				
	Analyze	Solve	Verify	Maintain
Declarative	11	1	0	0
Procedural	4	5	3	0
Schematic	4	0	0	0
Strategic	2	6	1	0

Student 4 (Phoenix_H)				
	Analyze	Solve	Verify	Maintain
Declarative	10	2	1	0
Procedural	2	12	0	0
Schematic	0	0	3	1
Strategic	5	6	1	1

Student 5 (Jessie_H)				
	Analyze	Solve	Verify	Maintain
Declarative	7	1	1	0
Procedural	2	6	4	0
Schematic	4	1	3	1

2015

(Student 1, Logan_L)				
	Analyze	Solve	Verify	Maintain
Declarative	10	7	2	0
Procedural	0	13	0	1
Schematic	0	0	1	0
Strategic	1	6	1	3

(Student 2, Armani_H)				
	Analyze	Solve	Verify	Maintain
Declarative	3	2	4	0
Procedural	2	10	2	3
Schematic	2	0	2	1
Strategic	2	6	4	1

(Student 3, Justice_H)				
	Analyze	Solve	Verify	Maintain
Declarative	10	4	3	0
Procedural	3	12	3	1
Schematic	3	2	1	1
Strategic	2	6	3	0

(Student 4, Alexis_L)				
	Analyze	Solve	Verify	Maintain
Declarative	11	8	5	0
Procedural	0	6	1	0
Schematic	1	2	0	1
Strategic	1	9	2	0

(Student 5, Emerson_H)				
	Analyze	Solve	Verify	Maintain
Declarative	7	8	2	0
Procedural	1	6	2	0
Schematic	1	0	2	0

2016

Student 1 (Oakley_L(H))				
	Analyze	Solve	Verify	Maintain
Declarative	2	8	2	0
Procedural	2	2	0	0
Schematic	2	1	1	0
Strategic	2	2	2	1

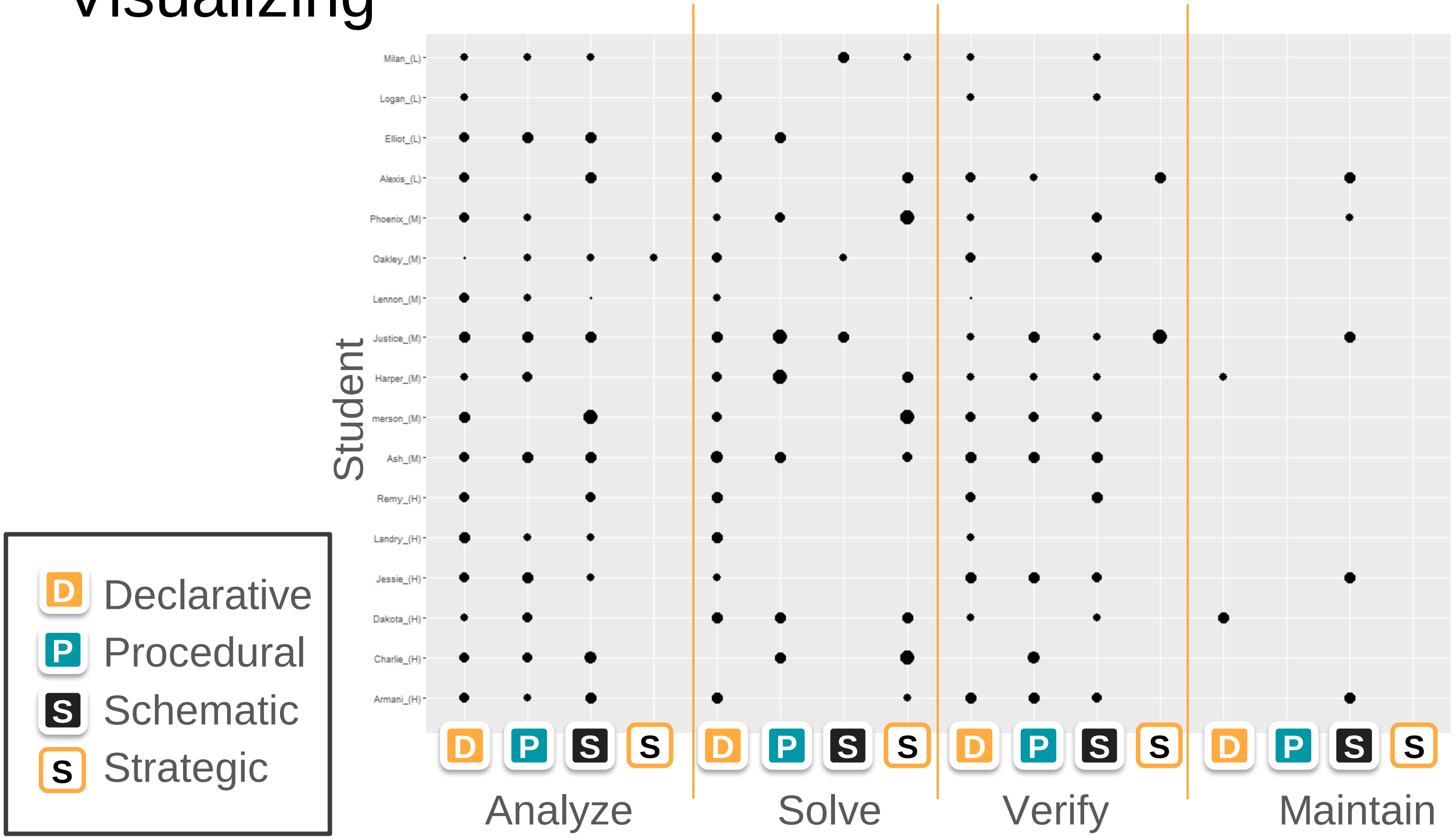
Student 2 (Dakota_L(H))				
	Analyze	Solve	Verify	Maintain
Declarative	10	7	4	1
Procedural	6	7	0	1
Schematic	0	0	1	0
Strategic	0	7	3	0

Student 3 (Harper_L(H))				
	Analyze	Solve	Verify	Maintain
Declarative	8	4	3	2
Procedural	4	9	7	0
Schematic	0	0	2	0
Strategic	4	6	6	0

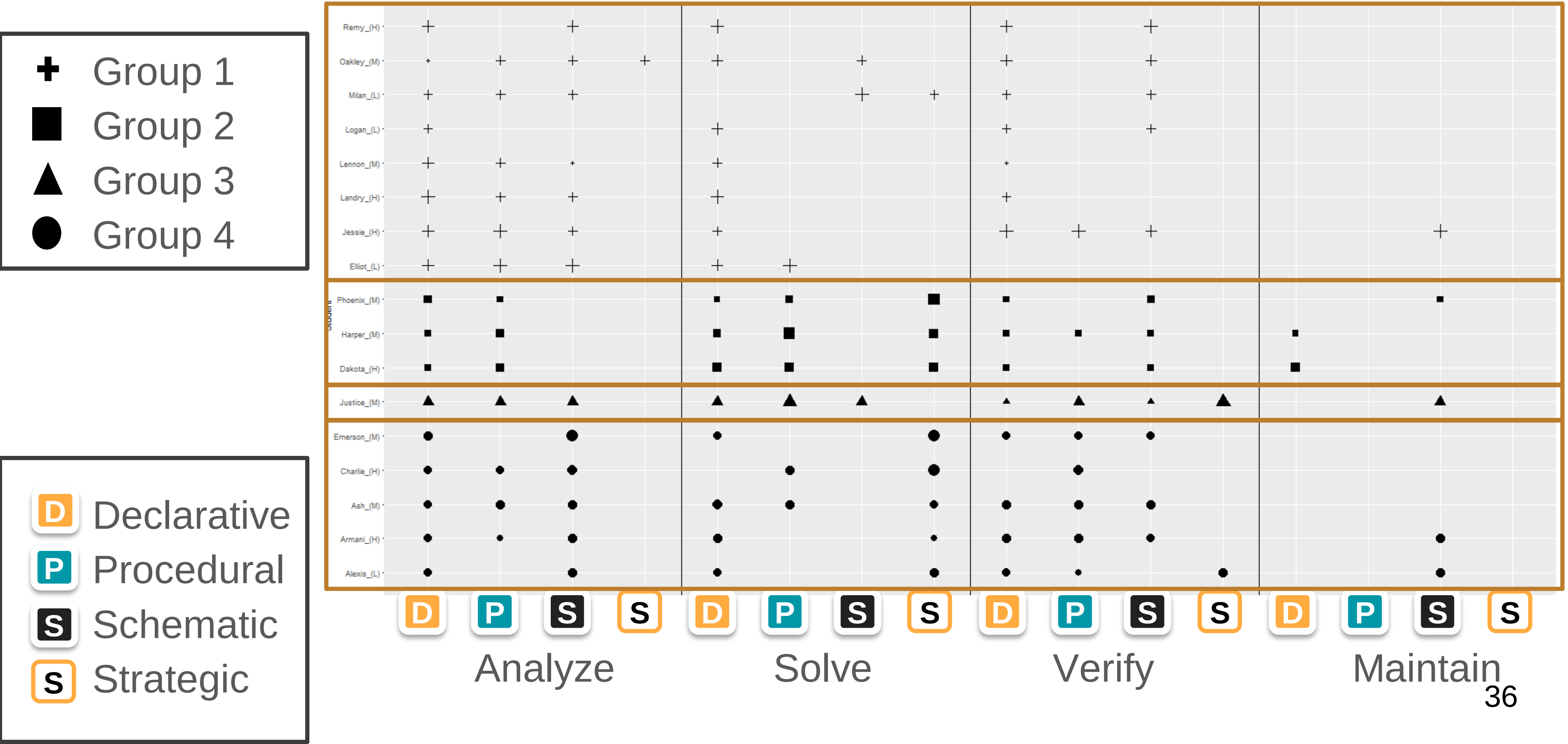
Student 4 (Landry_H(H))				
	Analyze	Solve	Verify	Maintain
Declarative	11	5	3	0
Procedural	1	3	0	0
Schematic	3	0	0	0
Strategic	1	2	2	1

Student 5 (Ash_H(H))				
	Analyze	Solve	Verify	Maintain
Declarative	7	6	4	0
Procedural	8	6	3	1
Schematic	1	0	1	0

Visualizing



Visualizing – Clustered by counts (K-means)



Exploration | Categorization | Grouping | Validation

Artifact Analysis

Identify Learner Profiles of habits of mind via document analysis

Analysis of think-aloud data

Habits of Mind:

- A person's outlook on knowledge and learning and ways of thinking and acting.
- Mathematical and logical skills that are essential tools for both formal and informal learning and for a lifetime of participation in society as a whole.

Goal: Identify students' different ways of thinking when explaining phenomena in the domain of signal processing

Elicitation of Student Performance

Definition of Discrete Time Fourier Transform (DTFT)

$$X(\omega) = \sum_{k=-\infty}^{\infty} x[n]e^{-j\omega k}$$

index of signal and summation do not match shows lack of mathematical rigor and critical response skills (catching a mistake)

Tag : (B1, D1)

Definition of Inverse Discrete Time Fourier Transform (iDTFT)

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)e^{j\omega n}\,d\omega$$

$X(\omega)$ is seen to be periodic with a period of 2π to see this ω is replaced with $\omega + 2k\pi$ where k is an integer

$$X(\omega + 2k\pi) = \sum_{n=-\infty}^{\infty} x[n]e^{-j(\omega+2k\pi)n}$$

Using the multiplicative rule of exponential the ω and $2k\pi$ are split into two different exponential

$$X(\omega + 2k\pi) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}e^{2k\pi n}$$

error : missed a negative sign in the exponent small arithmetic mistake shows that rigor and computation are not perfect

given that n and k are integers k and so $e^{-j2k\pi n} = 1$ for all k, from Euler's identity and so

$$X(\omega + 2k\pi) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} = X(\omega)$$

Overall, the communication and explanation in this section is not perfect.

Tag : (A3, B3, C3)

so $X(\omega + 2k\pi) = X(\omega)$ for all ω

Qualitative Analysis – Traditional approach

“Values and attitudes” was operationalized as students’ critical views of their own work and that of their peers:

“I think specific outline is very helpful and make easy to follow the formula and graphs. Formulas and graphs are very clear to understand.”

“I think an important aspect that you did not include in your final answer is that the DTFT of a DT signal must be periodic. Your answer must be "rep-ed" to denote it's periodicity. Otherwise your answer is only correct for o<=w<=2pi. The DTFT of x[n] is $rep_{2\pi}(2\pi\delta(\omega-\omega_0))$ *Overall color coating was very helpful, and the slecture was concise and clear.”*

Quantification of Qualitative Data

Description		Level of Performance			
Element	Definition	Below Basic 1	Basic 2	Proficient 3	Advanced 4
Computation and Estimation	Ability to choose an appropriate computation method and carry out the mathematical procedure accurately	Student selected an incorrect method and the solution was completely off.	Student selected a correct method but the solution was incorrect.	Student selected an appropriate method and the solution was correct. However, the student did not provide a justification for the method based on the circumstances, or the justification was inadequate.	Student selected an appropriate method, the student provided a correct justification for the selection of the method based on the circumstances and the solution was correct.
Mathematical Rigor	Ability to handle mathematical rigor and remember details of a definition	Student was not at all rigorous in the involved mathematics.	Student displayed some rigor but there were major errors.	Student was very rigorous but made small errors.	Student was very rigorous and made no errors.

Interpretation

Model sequence of scores (tag) of each student as random process

For simplicity, assume successive scores are independent

Estimate probability of each score occurring

$$P_{ij} = \text{Prob}(\text{element}_i, \text{level}_j), \; i=1,2,3,4,5, \;\; j=1,2,3,4$$

Count the total number of scores: N

Parameters of random process used as features

Each student represented by vector in R^{21}

$$(P_{11}, P_{12}, P_{13}, \ldots, P_{45}, N)$$

Half of students used as “Observation Data”

- Remaining half used later as “Validation data”

Cluster features of students in Observation Data using random approach called N-TARP

- Results in many different binary clusterings

Each binary clustering gives a label for each student

- 1 if student belongs to first cluster
- 2 if student belongs to second cluster

Concatenate the cluster labels and add grade to obtain a profile for each student

E.g., student 1 represented by (1,1,2,1,1,1,2,2,2,2,1,1,1,1,1, B)

Use Validation Data to corroborate each binary clustering found

Half students in validation data

Used permutation test to check statistical significance of difference between clusters

Permutation test appropriate because data is small

Analyze differences in grades between students in different clusters

Discuss the findings

Cluster 1:

Habits developing

Element/Level	Below Basic	Basic	Proficient	Advanced
Computation	0	1.667	3.3333	4.1667
Rigor	3.3333	4.1667	5.0000	1.6667
Communication	0	5.0000	4.1667	1.6667
Critical Response	2.500	1.6667	3.3333	3.3333
Values	3.3333	41.6667	9.1667	0.8333

Cluster 2:

Habits developed

Element/Level	Below Basic	Basic	Proficient	Advanced
Computation	0	0.7812	0.3906	9.3750
Rigor	0.7812	1.9531	6.2500	5.0781
Communication	0.3906	1.9531	7.0312	5.0781
Critical Response	0.7812	0.7812	4.2969	6.6406
Values	0	27.7344	12.1094	8.5938

Grade	All Students	Cluster 1: "Habits Developing"	Cluster 2: "Habits Developed"
A (4.0)	5	0	5
B (3.0)	10	2	8
C (2.0)	8	4	4
D (1.0)	2	2	0
F (0.0)	2	2	0
Mean Grade	2.5185	1.600	3.0588
Standard Deviation	1.1222	1.0750	0.7475

Exploration | Categorization | Grouping | Validation

Survey Data

Visualize and group categorical data of students' preferences

Student Preferences

Indicate your preferences about the following programs

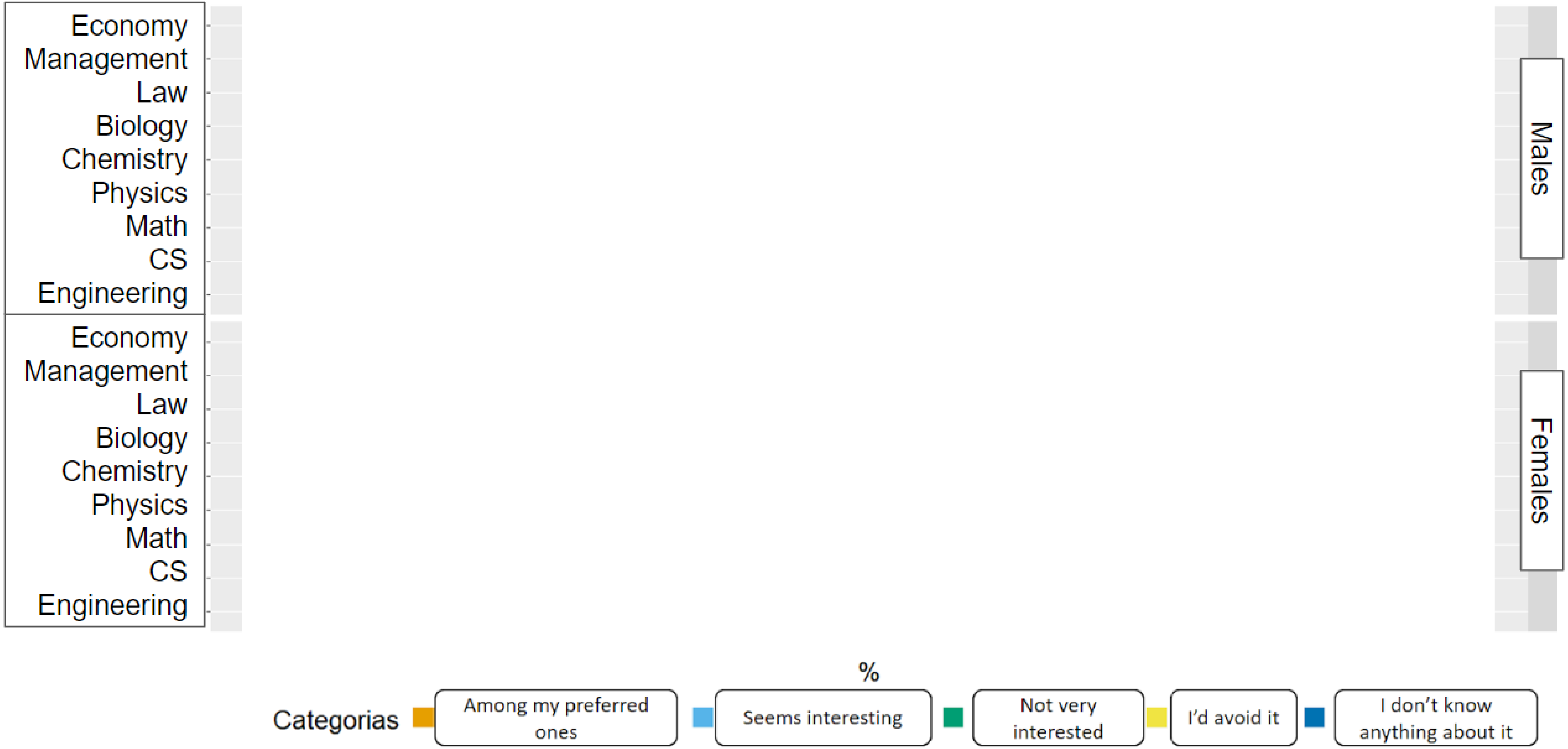
Economy, Management, Law, Biology, Chemistry, Physics, Math, CS, Engineering

[Among my preferred ones; Seems interesting; Not very interested; I’d avoid it;

I don’t know anything about it]

Goal: Identify student perceptions and preferences about different undergraduate programs

Student Preferences of Undergraduate Programs



Multiple Correspondence Analysis
+
Hierarchical Clustering
+
Chi-Square

- Those who avoid everything (2574, **25.4%**)

  Girls > Boys
- Those who don’t know any program (435, **4.3%**)

  Girls < Boys
- Those interested in everything (6035, **59.5%**)

  Girls <
- Those who choose not to answer (1101, **10.9%**)

  Girls < Boys

Exploration | Categorization | Grouping | Validation

Open-Ended Responses

Semi-automatic categorization of open-ended responses

Open-Ended Responses

Thermal and Transport Science Concept Inventory

Multiple choice assessment instrument that assesses the topics of “diffusion”, “heat transfer”, and “microfluidics”

For each of the TTCI multiple choice questions, an open-ended question prompted students to explain their rationale for choosing a particular response item.

Goal: Identifying student conceptual change from their open-ended responses

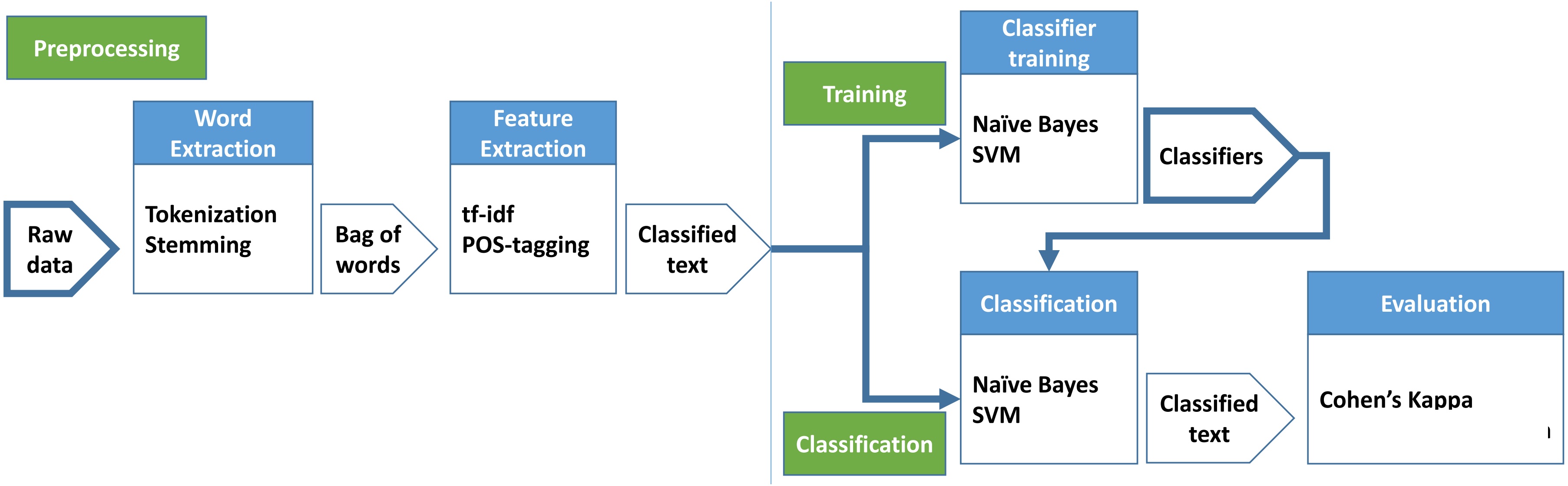
Traditional Approach

Thematic Analysis: Identifying categories and codes to identify patterns (themes) from the data

Content Analysis: Counting frequency of words or phrases

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, *33*(4), 445-459.

Data Analysis Process



Results

	Precision	Recall	F1	kappa
2-way				
Naïve Bayes classifier	0.7250	0.5713	0.6331	0.5346
SVM classifier	0.9095	0.6452	0.7496	0.6865
3-way				
Naïve Bayes classifier	0.6924	0.6904	0.6804	0.4545
SVM classifier	0.7408	0.7298	0.7199	0.5131
4-way				
Naïve Bayes classifier	0.6088	0.6145	0.6029	0.3523
SVM classifier	0.6414	0.6494	0.6340	0.4103

Results

	kappa
2-way	
Naïve Bayes classifier	0.5346
SVM classifier	0.6865
3-way	
Naïve Bayes classifier	0.4545
SVM classifier	0.5131
4-way	
Naïve Bayes classifier	0.3523
SVM classifier	0.4103

Let’s try it out

Let’s try it out

R Programming - Handout

1. We brought 6 computers - Create groups and distribute among the computers

2. Check out the handout

1. If you don’t have any experience with R, start by: ***IntroductionToR.R***

2. Plots: ***IntroToPLots.R***

3. Clustering: ***ClusteringPlusPlots.R***

Overall Conclusions and Next Steps

Computation can be used to **explore** complex educational data, as well as **identify, characterize and validate groups**

Different visualization, clustering, and validation techniques should be used depending on the data

The interpretation and connection with theory is responsibility of the researcher

R code available at www.github.com/cvieiram

How can you use this in your own research?

Acknowledgements

Prof. Alejandra J. Magana

Hayden Fennell

Prof. Mireille (Mimi) Boutin

Tarun Yellamraju

Dr. Camilo Vieira

Natalia Hernandez

This work has been partially supported by the NSF grant awards **EEC#1826099**

About Us

Dr. Camilo Vieira

Assistant Professor at Universida del Norte (Colombia)

cvieira@uninorte.edu.co

Prof. Alejandra J. Magana

Professor of the Computer and Information Technology Department at Purdue

admagana@purdue.edu

Prof. Mireille "Mimi" Boutin

Associate professor in the School of Electrical and Computer Engineering at Purdue

mboutin@purdue.edu

Thank you!

Questions?

Camilo Vieira, PhD
cvieira@uninorte.edu.co / @cvieiram

Prof. Alejandra J. Magana
admagana@purdue.edu
Prof. Mimi Boutin
mboutin@purdue.edu