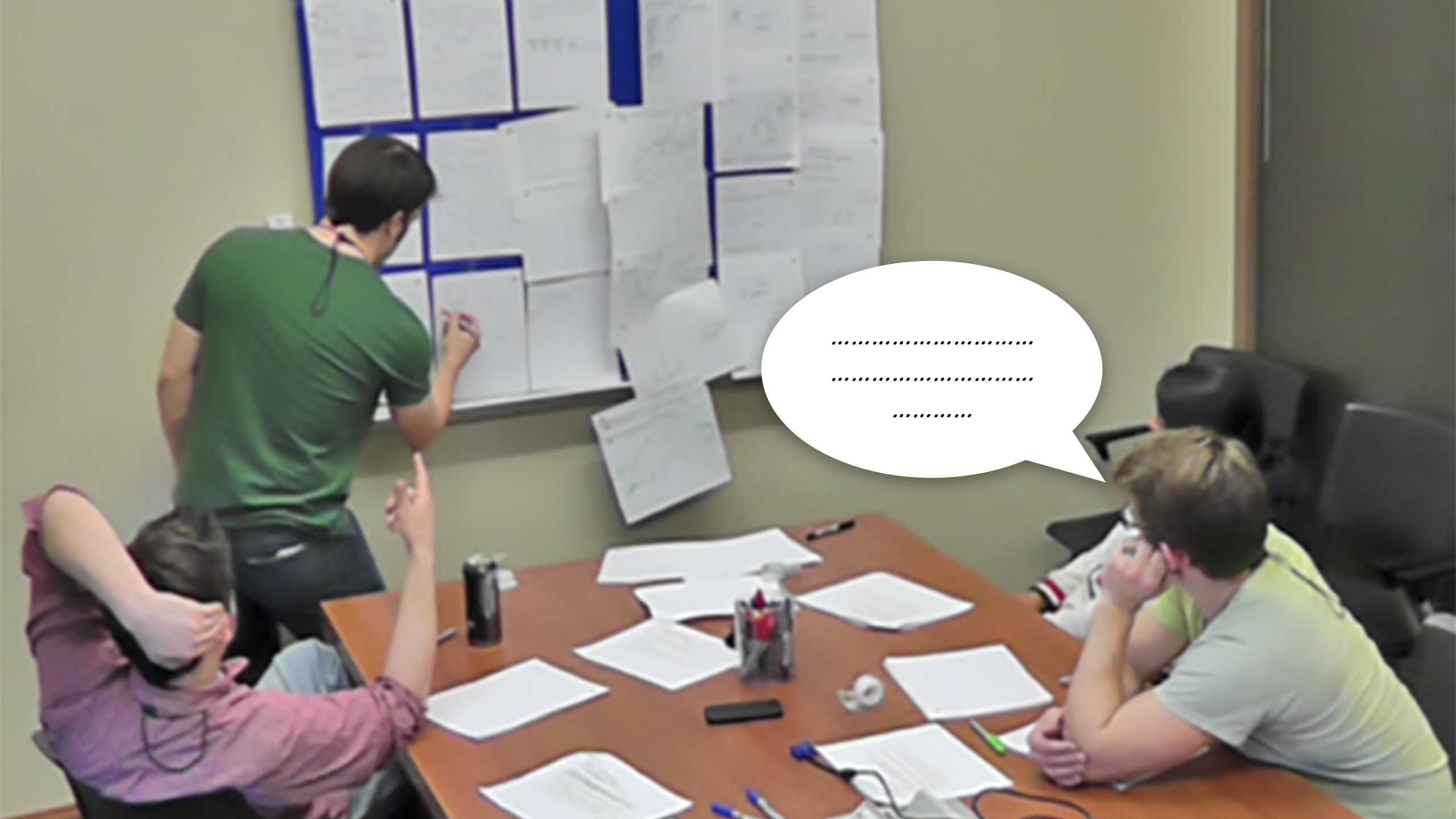
Visual Analysis of Unstructured Text Data

(and its application to Maintenance Work Orders)

Senthil Chandrasegaran
Postdoctoral Scholar
ViDi Labs, University of California-Davis

UCDAVIS
COMPUTER SCIENCE

VISUALIZATION & INTERFACE
DESIGN INNOVATION



Maintenance Data

Source: University of Western Australia System Health Lab (Excavator dataset)

StartDate	Asset	Original Short text	PM Type	Cost
7/1/04	A	BUCKET WON'T OPEN	PM01	183.05
3/20/05	A	L/H BUCKET CYL LEAKING.	PM01	407.40
5/5/06	A	SWAP BUCKET	PM01	0.00
7/11/06	A	FIT BUCKET TOOTH	PM01	0.00
11/10/06	A	REFIT BUCKET TOOTH	PM01	1157.27
4/21/08	A	REPAIR LOOSE BUCKET CYLINDER	PM01	0.00
10/2/08	A	REPLACE 2 BUCKET TEETH	PM01	1256.62
7/12/01	B	REPLACE MISSING LIP TIPS	PM01	1281.36
7/19/01	B	REPLACE LIP	PM01	1251.52
10/1/01	B	rebuild lip #3	PM01	0.00
10/10/01	B	REPLACE BUCKET TEETH.	PM01	681.10
10/29/01	B	Repair RH120 lip at P&H	PM01	22611.06
11/6/01	B	replace missing tip	PM01	1239.29
11/13/01	B	replace missing tip& adaptor.	PM01	221.15
11/16/01	B	Replace centre and LH lip shrouds	PM01	2118.87
1/15/02	B	OIL LEAK ON BUCKET....	PM01	0.00
2/6/02	B	REPLACE ADAPTOR LOCK	PM01	346.47
2/22/02	B	Fit a lip shroud	PM01	56.59
2/27/02	B	Refit left lip shroud	PM01	627.85
3/8/02	B	refit adapter and tooth	PM01	1186.45
3/12/02	B	FIT NEW TOOTH	PM01	397.57
3/12/02	B	cut off broken wear package	PM01	0.00

Maintenance Data

Source: University of Western Australia System Health Lab (Excavator dataset)

StartDate	Asset	Original Short text	PM Type	Cost
7/1/04	A	BUCKET WON'T OPEN	PM01	183.05
3/20/05	A	L/H BUCKET CYL LEAKING.	PM01	407.40
5/5/06	A	SWAP BUCKET	PM01	0.00
7/11/06	A	FIT BUCKET TOOTH	PM01	0.00
11/10/06	A	REFIT BUCKET TOOTH	PM01	1157.27
4/21/08	A	REPAIR LOOSE BUCKET CYLINDER	PM01	0.00
10/2/08	A	REPLACE 2 BUCKET TEETH	PM01	1256.62
7/12/01	B	REPLACE MISSING LIP TIPS	PM01	1281.36
7/19/01	B	REPLACE LIP	PM01	1251.52
10/1/01	B	rebuild lip #3	PM01	0.00
10/10/01	B	REPLACE BUCKET TEETH.	PM01	681.10
10/29/01	B	Repair RH120 lip at P&H	PM01	22611.06
11/6/01	B	replace missing tip	PM01	1239.29
11/13/01	B	replace missing tip& adaptor.	PM01	221.15
11/16/01	B	Replace centre and LH lip shrouds	PM01	2118.87
1/15/02	B	OIL LEAK ON BUCKET....	PM01	0.00
2/6/02	B	REPLACE ADAPTOR LOCK	PM01	346.47
2/22/02	B	Fit a lip shroud	PM01	56.59
2/27/02	B	Refit left lip shroud	PM01	627.85
3/8/02	B	refit adapter and tooth	PM01	1186.45
3/12/02	B	FIT NEW TOOTH	PM01	397.57
3/12/02	B	cut off broken wear package	PM01	0.00

Importance of Visual Analysis

Human in the Loop: Make use of analyst's tacit knowledge

Visual pattern identification: Visual representations allow detections of patterns & anomalies

Text logs → NLP + Text Visualization

Providing **contextual information** when visualizing text

Visualizing **intrinsic & extraneous measures** to identify patterns/anomalies

Integrating **machine learning & visualization** for analysis/monitoring

Text logs → NLP + Text Visualization

Providing **contextual information** when visualizing text

Visualizing **intrinsic & extraneous measures** to identify patterns/anomalies

Integrating **machine learning & visualization** for analysis/monitoring

USA
1
...MAYBE USE A SERVO MOTOR ON THE TURRET?

VizScribe: Visualizing Text + Context

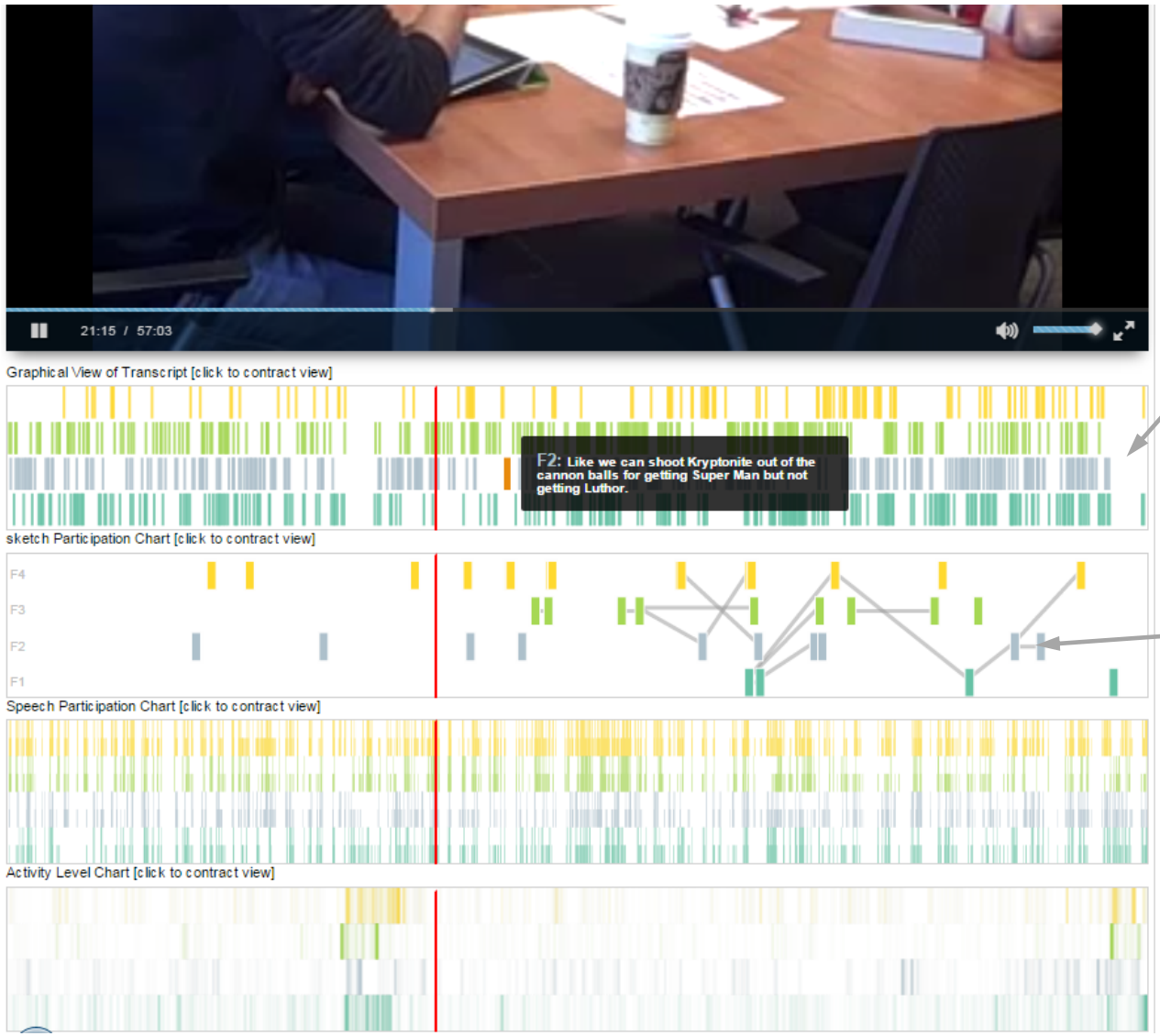
video

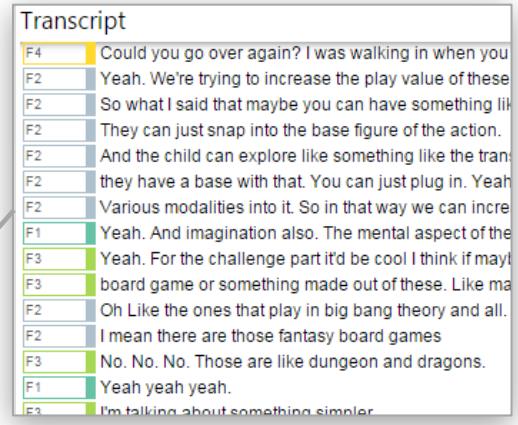
transcript

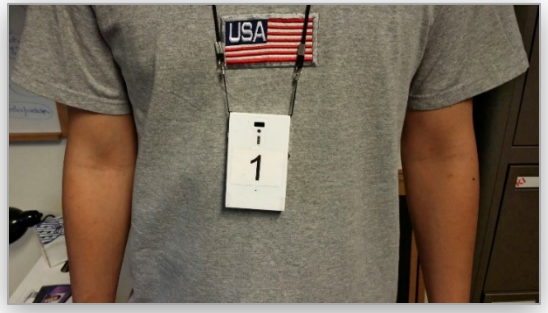
sketch log

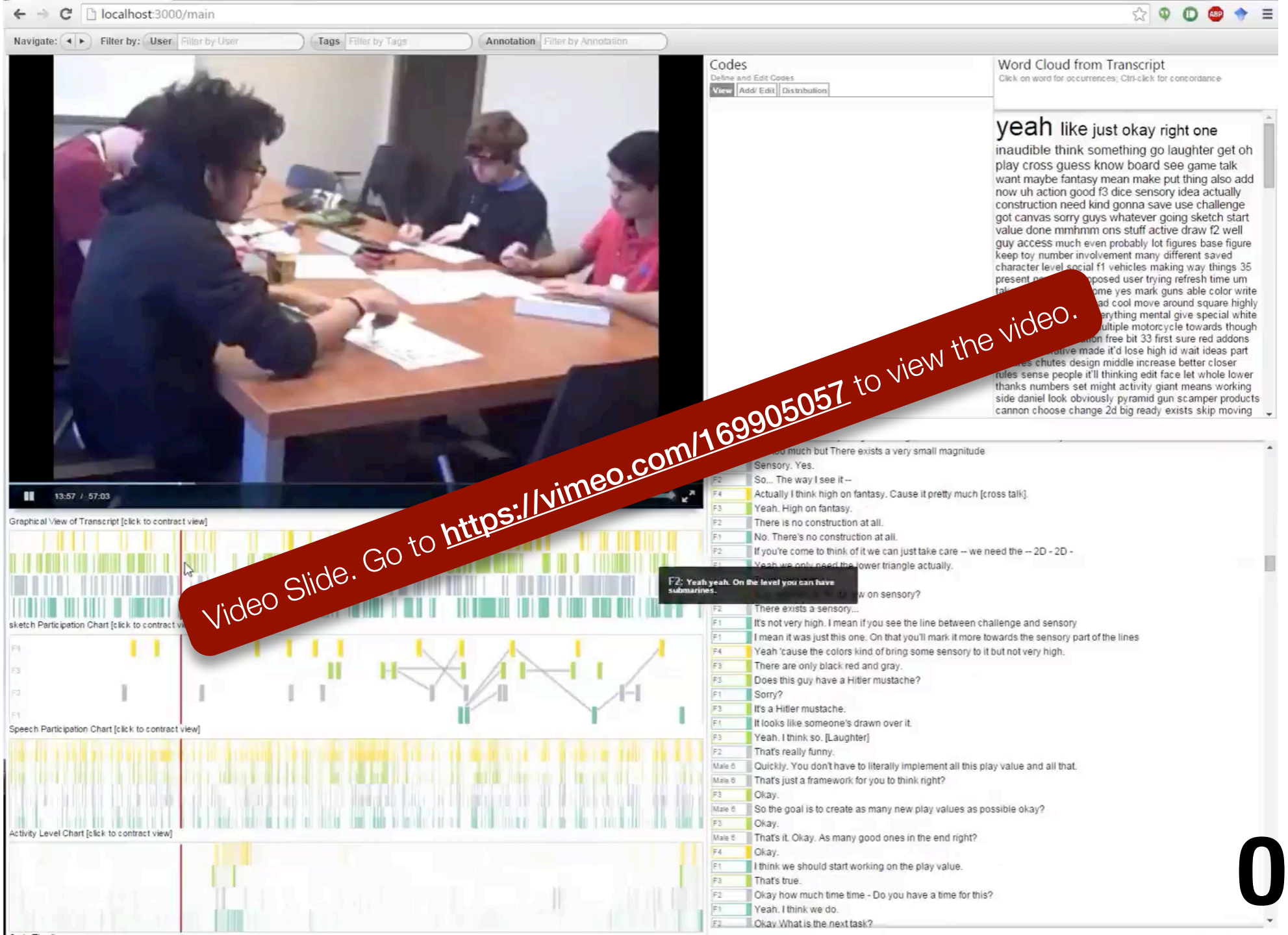
speech participation

activity data



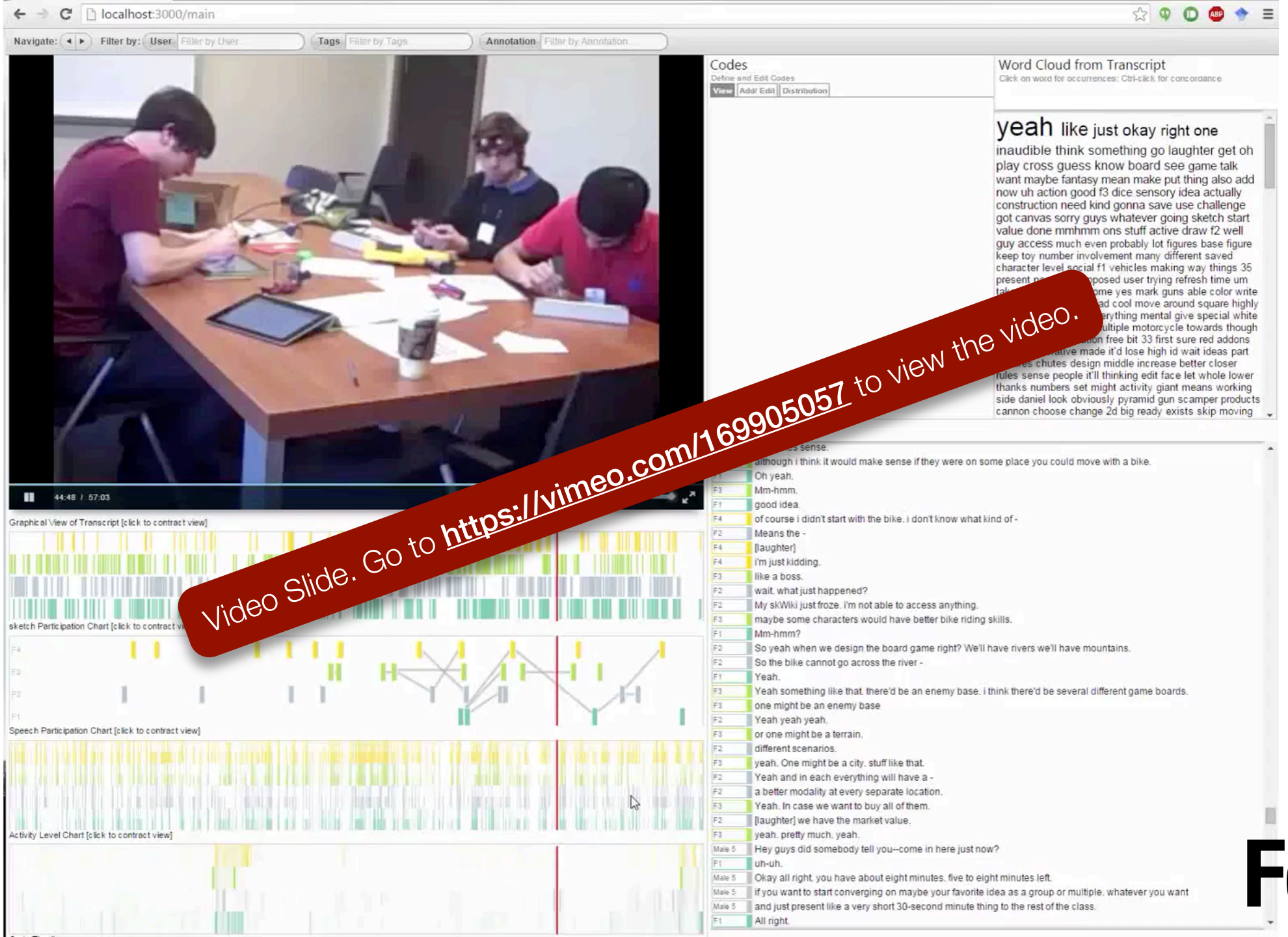






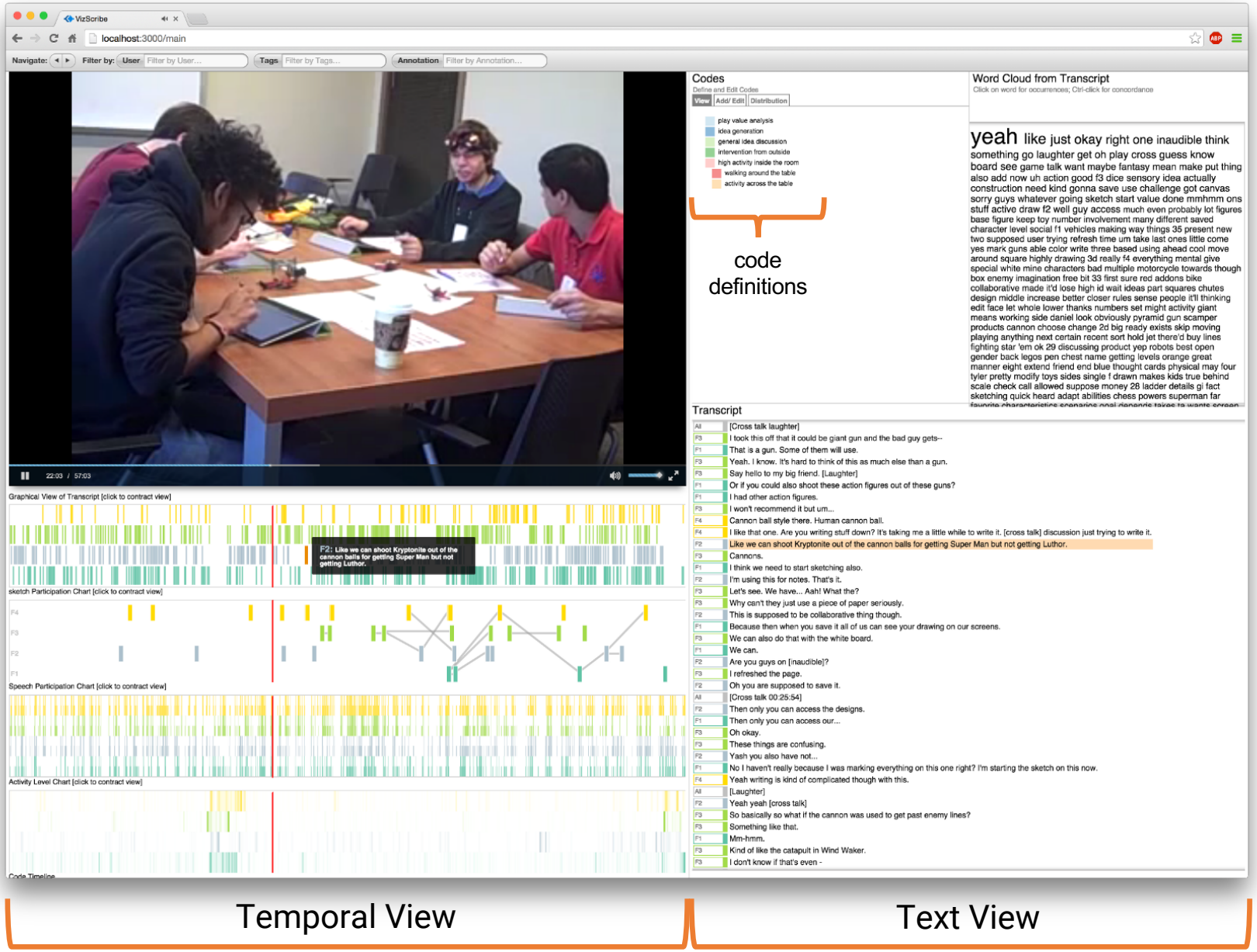
Video Slide. Go to https://vimeo.com/169905057 to view the video.

Orient


localhost:3000/main
Navigate:
Filter by:
User
Filter by User
Tags
Filter by Tags
Annotation
Filter by Annotation
Codes
Define and Edit Codes
View
Add/ Edit
Distribution
Word Cloud from Transcript
Click on word for occurrences; Ctrl-click for concordance
yeah like just okay right one
inaudible think something go laughter get oh
play cross guess know board see game talk
want maybe fantasy mean make put thing also add
now uh action good f3 dice sensory idea actually
construction need kind gonna save use challenge
got canvas sorry guys whatever going sketch start
value done mmhmm ons stuff active draw f2 well
guy access much even probably lot figures base figure
keep toy number involvement many different saved
character level social f1 vehicles making way things 35
44:48 / 57:03
Graphical View of Transcript [click to contract view]
sketch Participation Chart [click to contract v
Speech Participation Chart [click to contract view]
Activity Level Chart [click to contract view]
Video Slide. Go to https://vimeo.com/169905057 to view the video.
Oh yeah.
Mm-hmm.
good idea.
of course i didn't start with the bike. i don't know what kind of -
Means the -
[laughter]
i'm just kidding.
like a boss.
wait. what just happened?
My skWiki just froze. i'm not able to access anything.
maybe some characters would have better bike riding skills.
Mm-hmm?
So yeah when we design the board game right? We'll have rivers we'll have mountains.
So the bike cannot go across the river -
Yeah.
Yeah something like that. there'd be an enemy base. i think there'd be several different game boards.
one might be an enemy base
Yeah yeah yeah.
or one might be a terrain.
different scenarios.
yeah. One might be a city. stuff like that.
Yeah and in each everything will have a -
a better modality at every separate location.
Yeah. In case we want to buy all of them.
[laughter] we have the market value.
yeah. pretty much. yeah.
Hey guys did somebody tell you--come in here just now?
uh-uh.
Okay all right. you have about eight minutes. five to eight minutes left.
if you want to start converging on maybe your favorite idea as a group or multiple. whatever you want
and just present like a very short 30-second minute thing to the rest of the class.
All right.

Focus

Extension to Maintenance logs



Text logs → NLP + Text Visualization

Providing **contextual information** when visualizing text

Visualizing **intrinsic & extraneous measures** to identify patterns/anomalies

Integrating **machine learning & visualization** for analysis/monitoring

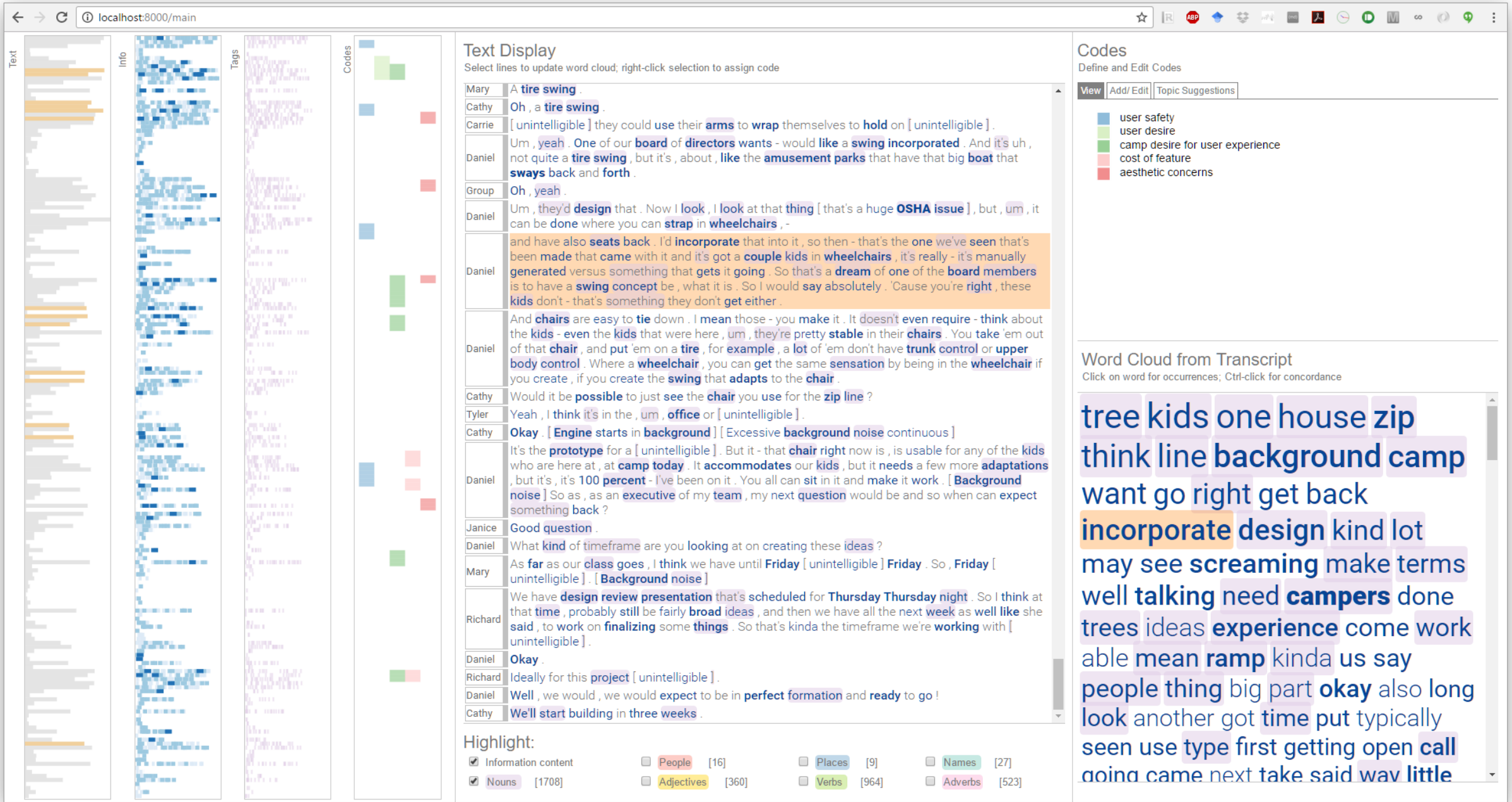
Qualitative text analysis

Open Coding	**Identify concepts** of interest in qualitative data; name & tag them
Axial Coding	**Identify relationships** between concepts through inductive reasoning

Help identify concepts

Help compare concepts & identify relationships

Prototype Interface



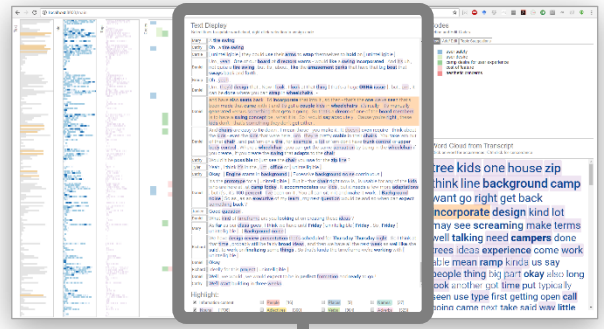
Identifying Concepts

Um , the maj - , the **majority** of the **campers** are gonna be with **people** [unintel
campers will have all the **emergency** equipment they **need right** there . So if th
anything that **happens** , we can **address** it wherever they are . **Accessibility** to
think we'll we do - you do that now . That's the be **caution** - **cautious** with that
my **eyes** , that's , that's a **low** , **low priority** .

Sure .

One **approach** to that **may** be , though , **accommodating** able - **bodied people**
might be that **similar** to what you do at a **main** entrance . The **ramp** is **primary**
also have another **route** that **might** be **steps** or , or a **steeper ramp** that's **sort**
express route for **accommodating** people from **two** different , um , **paths** . Tha
be out of the **question** at all to have some **steps** or , or **extra** - **steep** -

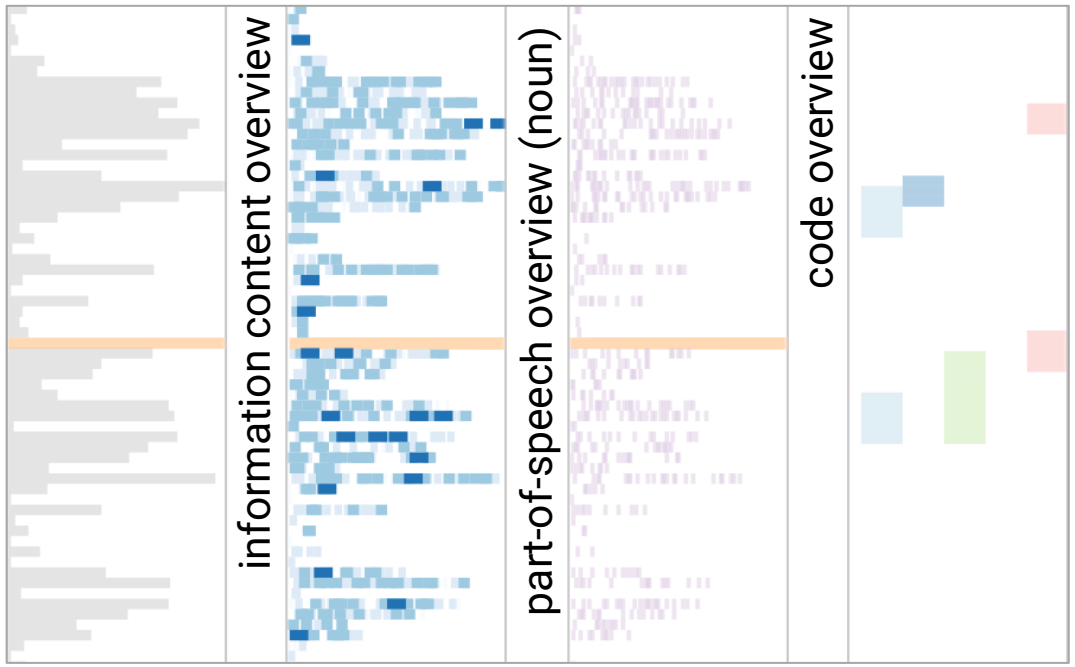
ramping from **multiple** places or at **least** a **single** location . Um , **one** of those
or uh , the **staff** that's actually **working** at the **zip line** to get up and down quick

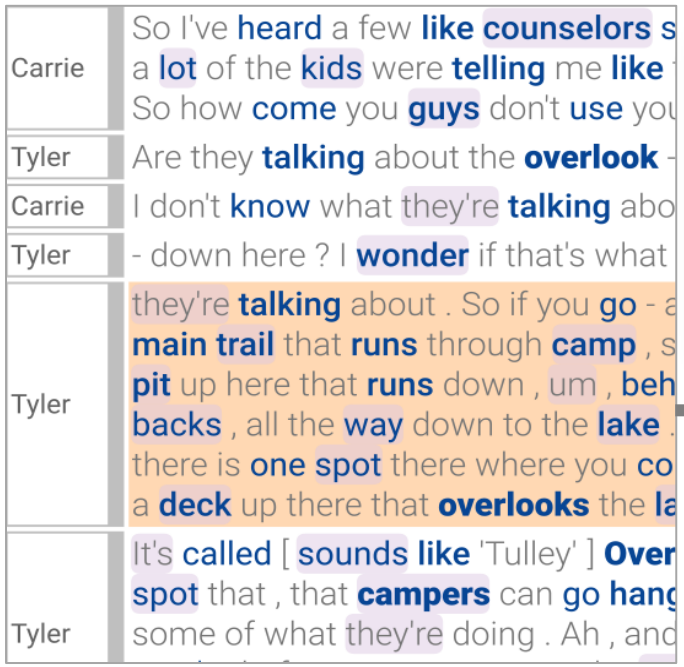


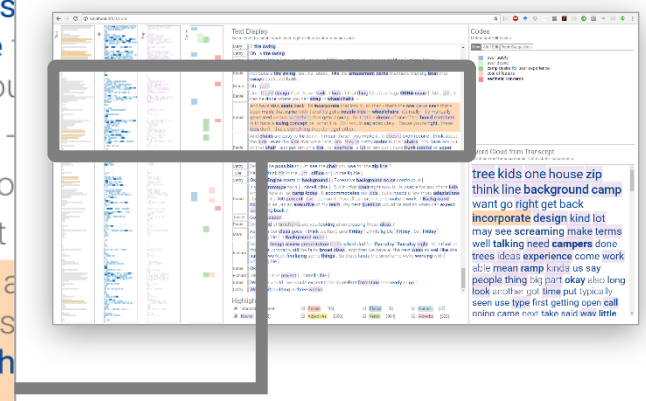
Skim formatting (filter by information content)

Highlight parts of speech (nouns, adjectives etc.)

Multiple Abstractions







localhost:8000/main
Text
Info
Tags
Codes
Text Display
Select lines to update word cloud; right-click selection to assign code
Tyler a few weeks ago . Right now , they have to sit down at the bottom of the zip line waiting to get geared up and before they can go up there and so [very loud machinery in background] , one , the ramp to get up there [unintelligible] which you kinda saw is a big flat black swinger arm and [four to one] pulley system to get them up to the top . So , the ramp would obviously improve that . I think like you said , just the novelty and the , the uniqueness factor of being up in the trees , um , can be something in and of itself for that .
Daniel Yeah , and , and then even incorporating the zip line because these kids , all campers like to see their friends go .
Tyler Yeah .
Daniel And it's that first push off the platform where it's like , 'Okay . I'm trusting that I'm gonna make it . ' Um , right now , they don't get that experience . They see them , they see the expression of the kid when they land , but it's that -
Daniel that descent is -
Tyler And , and to be with them and offer encouragement and -
Daniel Absolutely .
Tyler - you know the kids are up there when we ran it a few weeks ago , by themselves .
Daniel Right .
Tyler No one else - I mean Daniel was up there and one of our facilitator
um , there's no moral support up there with them in terms
that they're particularly close to who could encourage
Mary I have a question for you guys . Um , so in
useful thing we can give to you to
design that's kind of like
incorporate ? I mean
Daniel I could
Matt This is a big great space so design elements whether particularly interesting or innovative that [screaming in background] you might be able to dream up to we may not think of or the designers may not think of , um specifically with assisting these kids that , um , might not be common sense to us with creating lot of design over and over , but this is something unique that you can -
Matt wrap your minds around . But as far as getting into the , you know , the detail design or structural [unintelligible] or spatial relationships , um , [it could be generalizable] .
Cathy And none of us have that expertise . We have one key [unintelligible] , and that's it . The TA . [Engine starts in background] [But most of us are not even - unintelligible] .
Tyler Yeah , but obviously , you know -
Cathy Right .
Tyler - it's not , it's not a totally -
Cathy Yeah .
Tyler - unrealistic conception [to sketch either] because you guys , you know , the existing trees and the
Video Slide. Go to https://vimeo.com/194922904 to view the video.
Highlight:
Information content
Nouns [1708]
People [16]
Adjectives [360]
Places [9]
Verbs [964]
Names [27]
Adverbs [523]
Codes
Define and Edit Codes
View
Add/ Edit
Topic Suggestions
Word Cloud from Transcript
Click on word for occurrences; Ctrl-click for concordance
tree kids one know house zip think line background camp want things guys get go design incorporate back kind lot maybe may see campers terms make need talking well able ramp come done trees experience ideas mean work thing say us kinda people big okay part look got typically first long another use seen time put type also getting even vision said call came way sure take going little noise open next year obviously wheelchairs woods nature safety

localhost:8000/main
about four-foot level. Anything above that is outside of their normal element. We try to bring those elements to them, and we try to do that at The Camp as well. So we incorporate programming to the extreme, and that is what can we provide, ah, visibly challenged population of campers from adolescence, from child adolescence up to adults that they wouldn't typically be able to do.
Text Display
Select lines to update word cloud; right-click selection to assign code
Daniel
kids who come to The Camp require a lot more , ah , technological need maybe a wheelchair . So maybe indoor require ventilator assistance at nighttime . So these are typically kids that are not allowed to go to any other kinds of camps because of their medical condition . And as we came back to Bradford Woods three years ago and then Ohio for 20 years , um ,
Daniel
camps who can incorporate what The Camp brings to the table , we can handle the kind of kids we bring into our camp , and we see a facility can handle any type of camper just because of the physical demands of the environment , um , and then the physical demand to the camper . From an equipment perspective , it's [unintelligible] perspective . Um , i - , it is oftentimes that The Camp built new - established types of programs who want to do through the eyes of campers . Kids that you work - have seen today , this is - I've seen wheelchairs what there is what they see . While you have -
Daniel
some kids who are ambulatory , but a lot of the kids who are in wheelchairs their life's vision is from about four feet - about four - foot level . Anything above that is outside of their normal element . We try to bring those elements to them , and we try to do that at The Camp as well . So we incorporate programming to the extreme , and that is what can we provide , ah , visibly challenged population of campers from adolescence , from child adolescence up to adults that they wouldn't typically be able to do .
Daniel
And we've pushed the limits from , um , initially , um , there's - we , we get [tree to tree
that had ropes that we could , we can send up , ah , to all up to 25 - foot high
the ability to be up in a tree . Um , one confined to a wheelchair . A lot
you ask them whether they wanna do things inside or outside
wanna do things outside because they're [life is
Daniel
And that's their life . So we want it to be differ
what they want , um what they hope
people walk through] .
Female
Sorry , guys .
Um , swim
canoes that are not tippable , and so [unintelligible]
and outriggers . These were designed , um , kind of conceptually
and then we've created those . So literally , those canoes will not tip over
on ventilators in that canoe without the fear of them tipping . Um , and that
[they're missing] -
background] Ryan , Ryan .
For a long time [group of people walk through] we've talked about how we inco -
Daniel
how can we incorporate a zip line into camp activities . And if we were gonna do that , how could we incorporate more activities . So through some work with another camp out in Adirondack , ah , how to make , uh zip line adaptable for campers who are in wheelchairs or [rigid] , um , we've found a way to do it and work with the company that makes adaptive ski chairs out in Denver to create a chair that allows kids to [master it] . So we're pushing envelope . So the zip lining is great .
Daniel
We've incorporated it this year on the 100 - foot zip and that's opened the eyes to these kids and what they wanna do further . And so now we wanna incorporate it further . [We'll go further] always with the safety aspect in mind first . So now looking at the zip line going 400 feet down across the valley . And while that's all well and good , that's just a piece of the activity that we see and how we incorporate nature into that , the experience , and that's where the tree house came into play .
Daniel
When we talk about what's that tree house incorporate , it's how do we get kids up so that it's not just the zip line experience , it's the whole getting up and into the canopy of the tree to be a part of nature . And what does that look like for a camper um prospectively ? So getting their ideas is important . That's how we
Highlight:
Information content
People [16]
Places [9]
Names [27]
Nouns [1708]
Adjectives [360]
Verbs [964]
Adverbs [523]
Codes
Define and Edit Codes
View
Add/ Edit
Topic Suggestions
Word Cloud from Transcript
Click on word for occurrences; Ctrl-click for concordance
kids camp tree want outside lot
physical life try inside things camps times see
adolescence incorporate handle four years require
camper typically seen perspective wheelchair
campers maybe foot wheelchairs bring walk people group
limit beyond push hope talk different incorporated say ask confined one
ability never high send ropes climb get initially limits pushed adults
population challenged visibly provide extreme programming well elements
element normal level feet vision ambulatory today work eyes programs types
established new built oftentimes equipment demand environment demands
type facility kind table brings ohio ago three woods bradford back came
condition medical kinds go allowed nighttime 20 ventilator assistance indoor
come 25 need technological child able
Video Slide. Go to https://vimeo.com/194922904 to view the video.

Text logs → NLP + Text Visualization

Providing **contextual information** when visualizing text

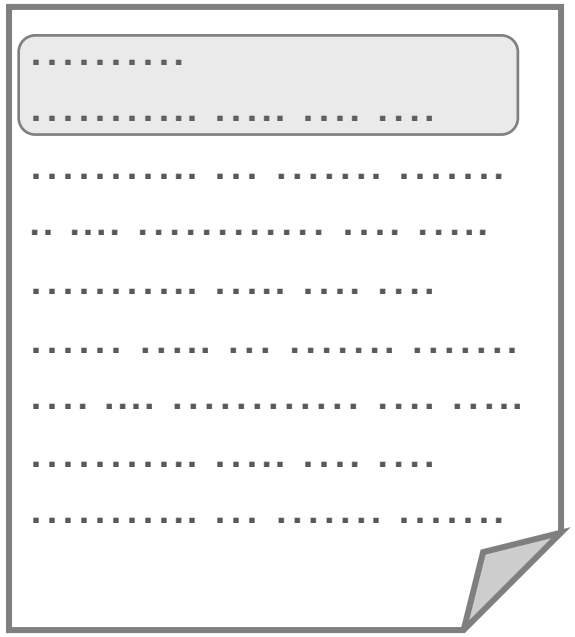
Visualizing **intrinsic & extraneous measures** to identify patterns/anomalies

Integrating **machine learning & visualization** for analysis/monitoring

TalkTraces: Monitoring Meetings

Meeting Transcript/

(or Maintenance logs)

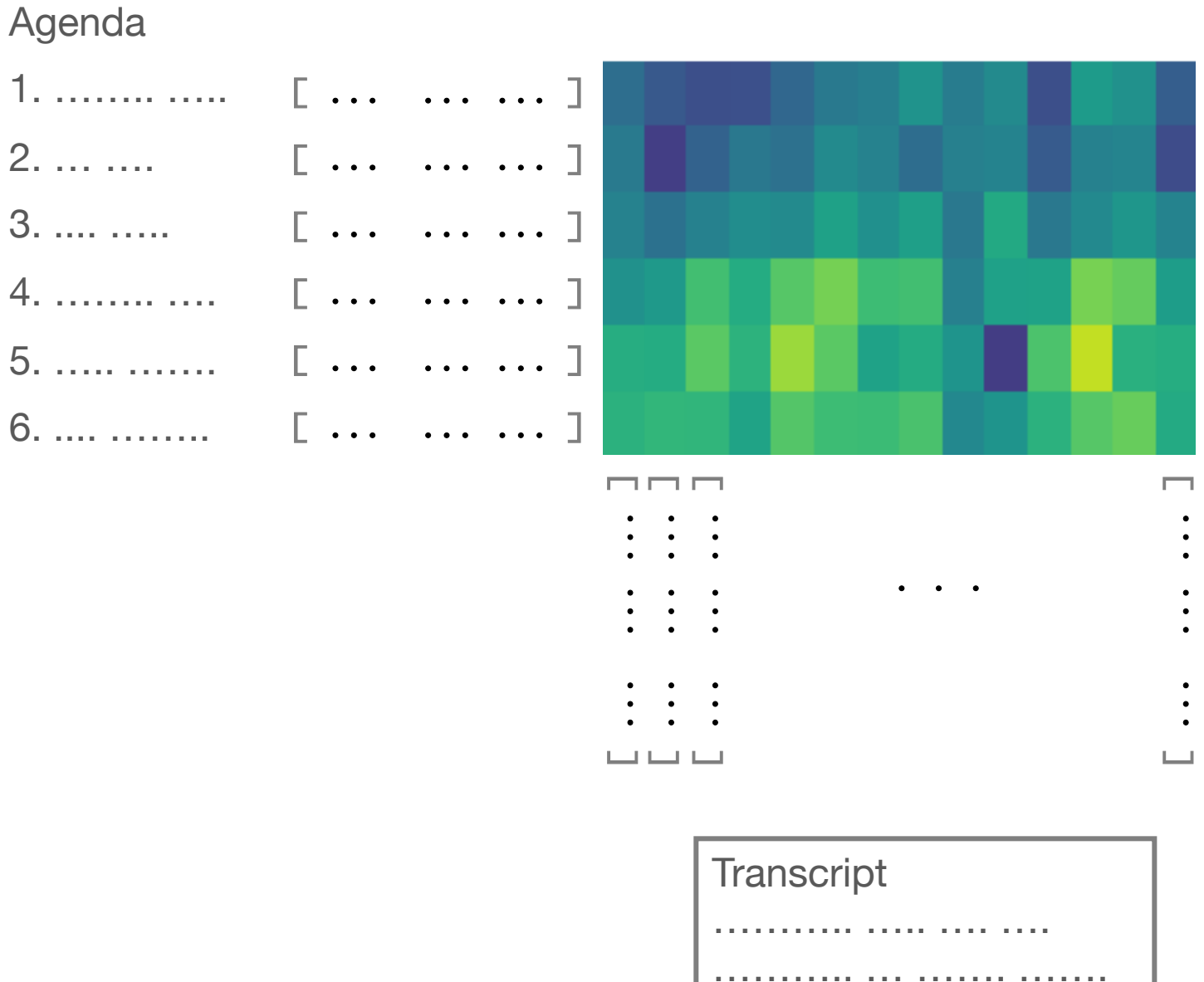


	T1	T2	T3	T4	
[	0.1	0.2	0.1	0.6	]
[	...	...	...	...	]
[	...	...	...	...	]
[	...	...	...	...	]
[	...	...	...	...	]

Providing Context

Using meeting agenda (or maint. log categories) to add **context**

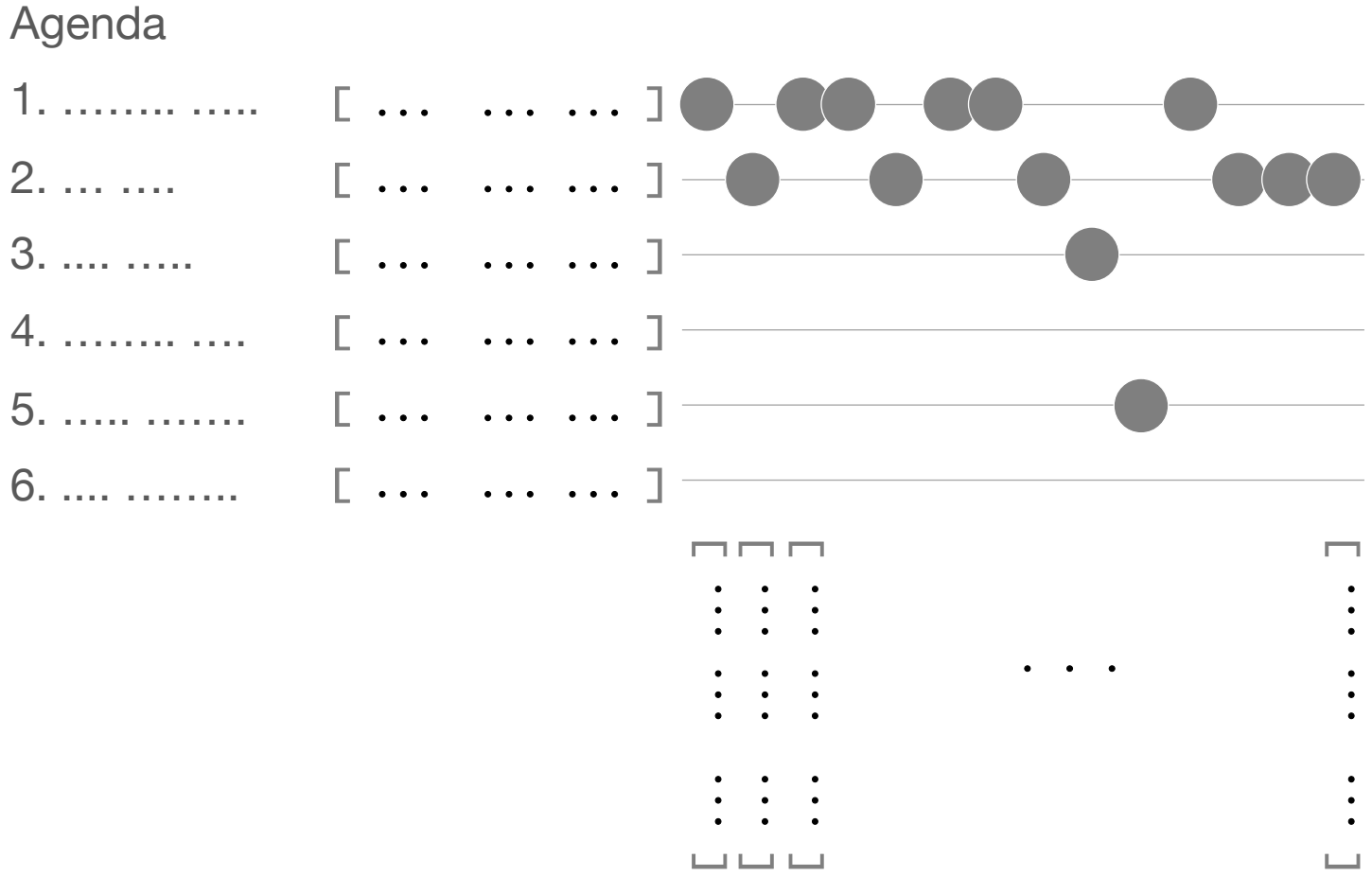
Compute relatedness between agenda/category & discussion/log using **word embedding**



Providing Context

Using meeting agenda (or maint. log categories) to add **context**

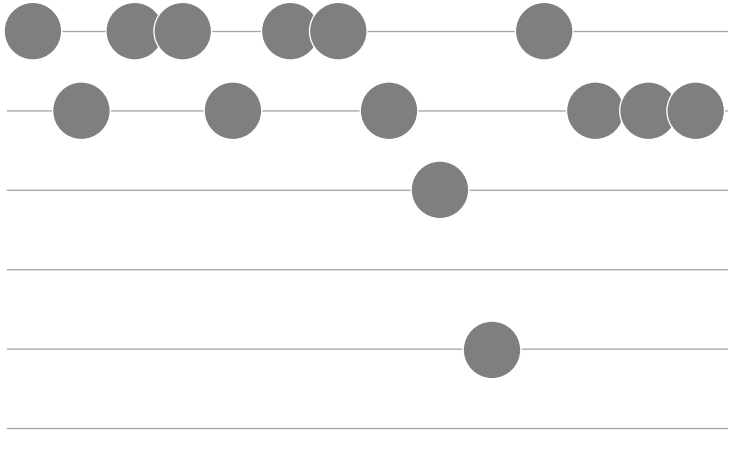
Compute relatedness between agenda/category & discussion/log using **word embedding**



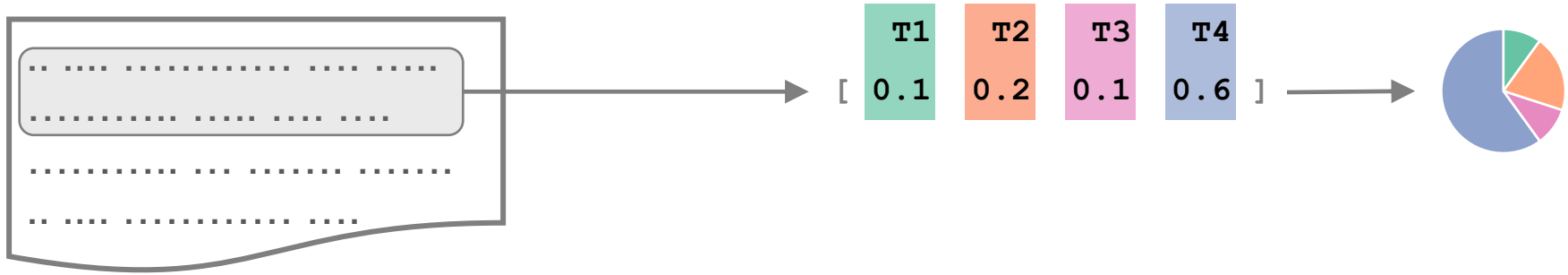
Providing Context

Agenda

1. …….. ….. [… … …]
2. … …. [… … …]
3. …. ….. [… … …]
4. …….. …. [… … …]
5. ….. ……. [… … …]
6. …. …….. [… … …]



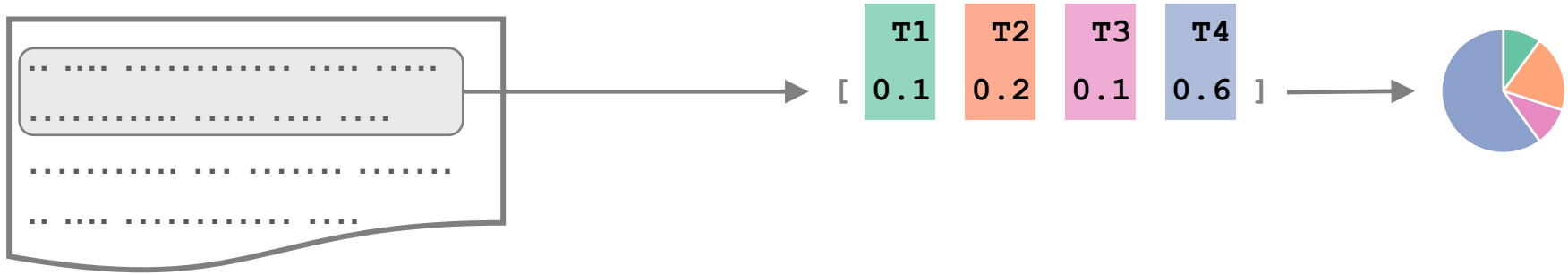
Providing Context

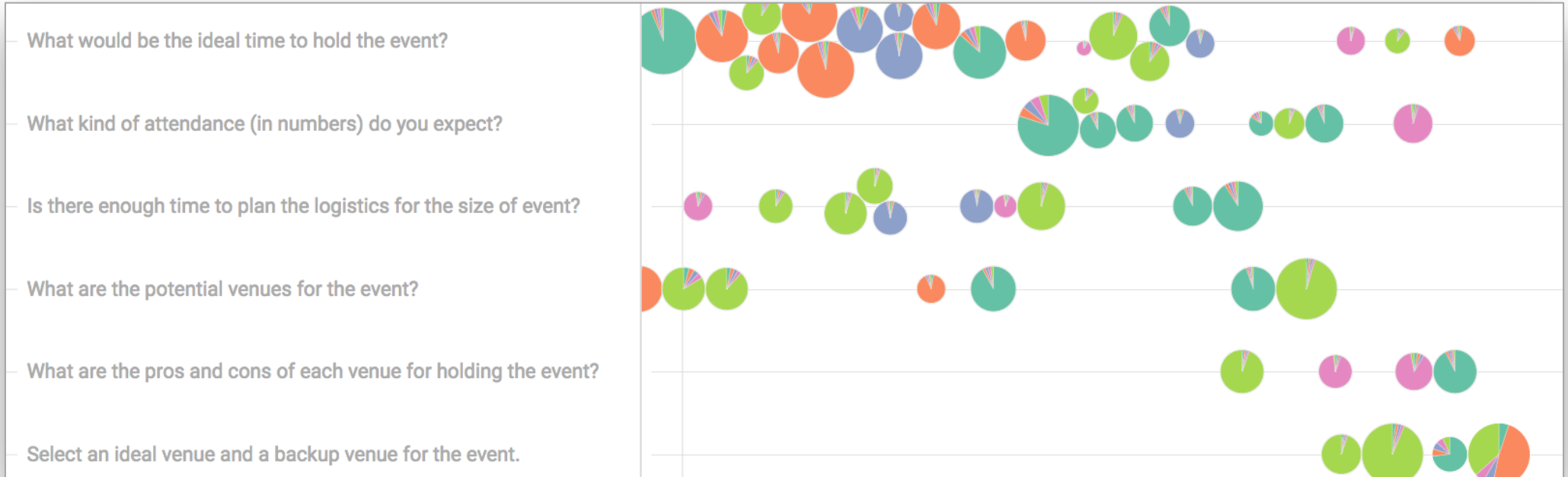


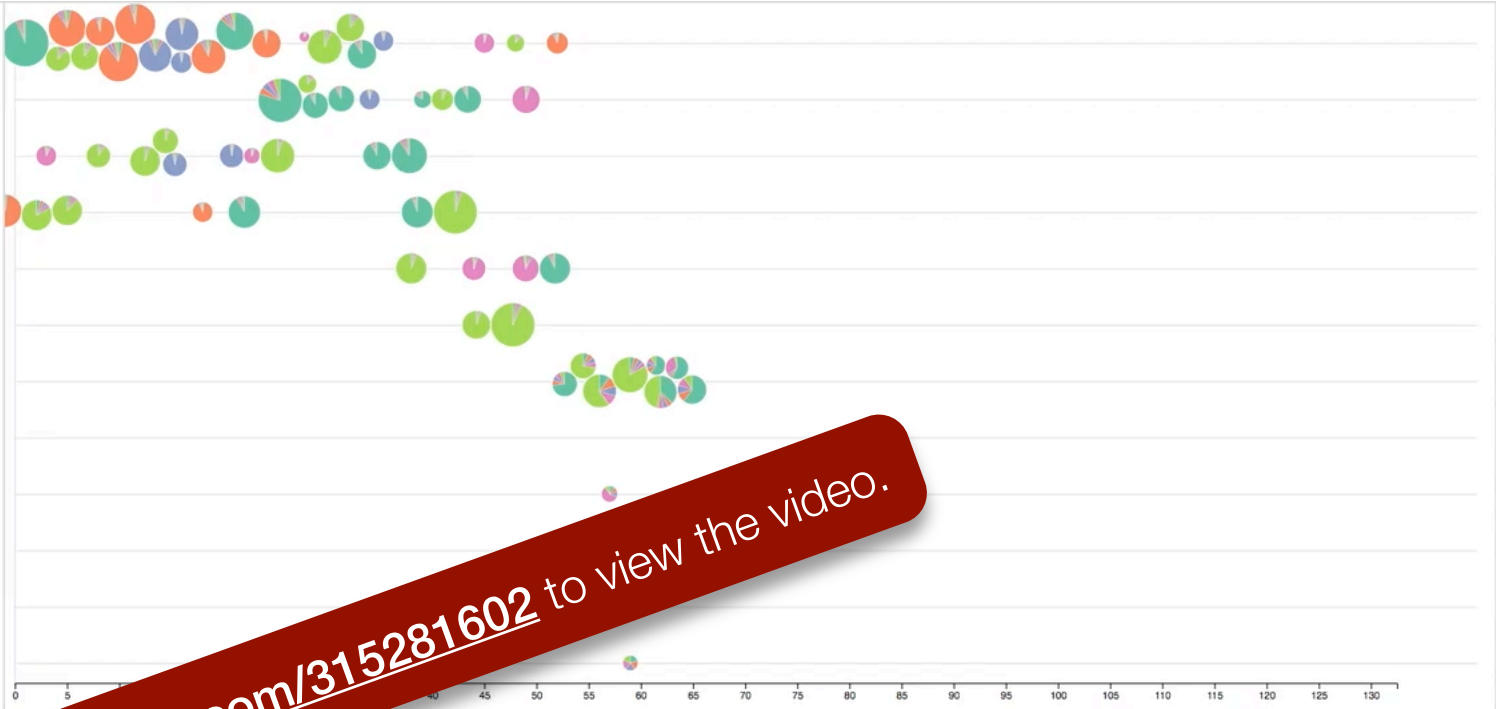
Agenda

1. …….. ….. [… … …]
2. … …. [… … …]
3. …. ….. [… … …]
4. …….. …. [… … …]
5. ….. ……. [… … …]
6. …. …….. [… … …]

Providing Context







Video Slide. Go to https://vimeo.com/315281602 to view the video.

T1	many lot people event picnic_day pro time show rsvp
T2	event kemper hold even time alumnus
T3	time saturday might alumnus building better though trying
T4	campus lunch forth people alumnus moved attendance tapping sound
T5	move people end show professor event matter time let

Show keywords on word cloud that are:

More Local — More Global

Relevance Weight: 0.40

time people idea weekend
even picnic_day best rock gun afternoon
might many lunch let multiple alumnus
campus date ideal sound week show good
better see move first saturday lot specific venue
combination pull con end coinciding night though

New keywords from current session:

need set run food stop agenda planning
morning require head justifying cilities discussing
catering health along line everybody presentation
man new projector permanent staff believe serve
poof serving hello reduce power pay covered

hh:mm:ss Otherwise this sounds good to me. So I guess one of the cons here would be this like limits the attendance that we can have. Right.
hh:mm:ss But pros is that it ticks like the demo and, Like talking them up boxes pretty well. If we're going for that. Okay so I think the main con is limits like capacity that we have.
hh:mm:ss Okay does anybody have any other comments? I think that covers all our topics. Okay well let's call it a day then.

13:15:25 hello everybody this morning were going to be discussing our next set of agendas about planning the event that we need
13:15:30 and how much money is going to require a set of heads
13:15:51 how many people
13:16:08 justifying cilities and then what do we actually need to run like we need catering so we can for health stuff
13:16:12 consideration
13:16:15 I think along those lines I think
13:16:17 stop
13:16:28 so we dont need specific stuff for the event because the people that are doing the presentations can also
13:16:44 be there in you really dont need a man to run the new projector and show the video of you dont even need permanent staff
13:16:53 also believe if we serve like poof and likes are serving food then we can reduce the main power
13:17:06 I think if we can get camper people to pay for the food like they do for other events so Im pretty sure if we can be covered
13:17:46 actually I dont think we need to make an appointment

actually I dont think we need to make an appointment

Playback speed : 8x

Lessons Learned

Visualizing **context** helps analysts use experiential knowledge in sensemaking

Tagging the data through the visualization helps **assign meaning**, detecting patterns/anomalies

Visualization can be used to help make ML results **interpretable**

Relevant Publications & Code

Chandrasegaran, S., Badam, S. K., Kisselburgh, L., Peppler, K., Elmqvist, N., & Ramani, K. (2017). VizScribe: A visual analytics approach to understand designer behavior. International Journal of Human-Computer Studies, 100, 66-80.

`https://github.com/senthilchandrasegaran/vizScribe`

Chandrasegaran, S., Badam, S. K., Kisselburgh, L., Ramani, K., & Elmqvist, N. (2017, June). Integrating visual analytics support for grounded theory practice in qualitative text analysis. In Computer Graphics Forum (Vol. 36, No. 3, pp. 201-212).

`https://github.com/senthilchandrasegaran/textplorer`

Chandrasegaran, S., Bryan, C., Shidara, H., Chuang, T-Y., and Ma, K.-L. (2019). TalkTraces: Real-Time Capture and Visualization of Verbal Content in Meetings. In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems, Paper 577, 14 pages.

`(TBD, please contact me for info)`

Contact

Email: senthil.kumaran@gmail.com | schandrasegaran@ucdavis.edu

Web: https://senthilchandrasegaran.github.io/