



Data Science for Smart Cities

CE88

Prof: Alexei Pozdnukhov
GSI: Madeleine Sheehan

115 McLaughlin Hall

alexeip@berkeley.edu
m.sheehan@berkeley.edu

CE88 in title



Today

Introduction and motivation: cities as complex systems.

Lecture 1. Introduction to urban systems. Inter-dependent infrastructures with human in the loop.

Lecture 2. Modeling principles. Causality and experiments in demand- and supply-side data analysis.

Agenda:

9:10 **Lecture 3. Spatio-temporal nature of urban data.**

9:50: Mini-lab setup

10:00 Break

10:10 Mini-lab

10:40 Lecture 4. Data flows in cities. Decision making feedback loops.

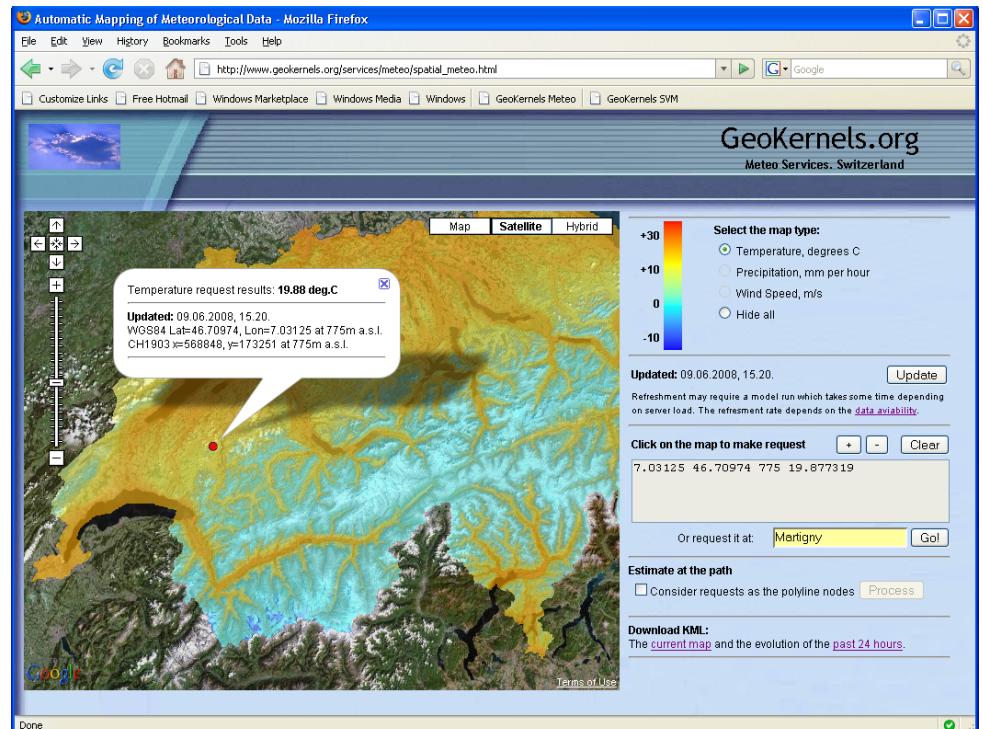
Spatio-temporal nature of urban data

- Data are often geo-referenced: there is a pair of coordinates* that links the data to a particular geographic location
- Data are often time-stamped: there is a moment in time or some temporal extent to which the data are relevant

Processes we study have distinct spatial and temporal properties

Data describing these processes have distinct spatial support

Conclusion 1. Space and time – i.e. maps and timelines
– provide a natural interface to visualize and explore urban data.



*) – such as latitude and longitude, or (x,y) in projected coordinate systems



Spatial data types

It is useful to use abstract data types that describe spatial extent of data.

Vector data provide a way to represent real world features. A vector feature has its shape represented with its geometry.

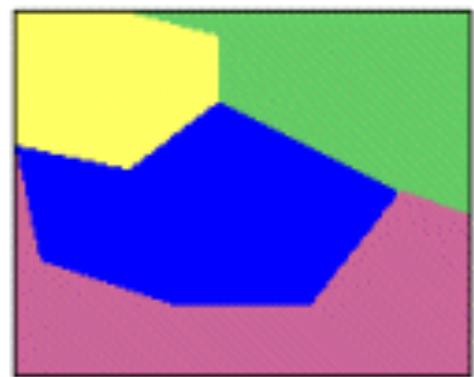
Points



Lines



Polygons



Vector data are good at

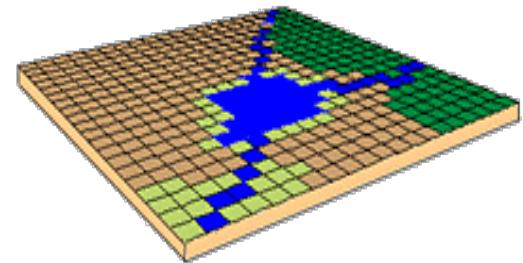
- accurately representing true shape and size
- representing non-continuous data (e.g., rivers, political boundaries, road lines, mountain peaks)
- creating aesthetically pleasing maps while conserving disk space



Spatial data types

Raster Data

Cell-based data representing digitized continuous space phenomena. Basically, a raster is a giant 2D table, where each element is assigned a specific numeric value. The meaning behind these values is specified by the user: they can represent elevations, temperature, etc.

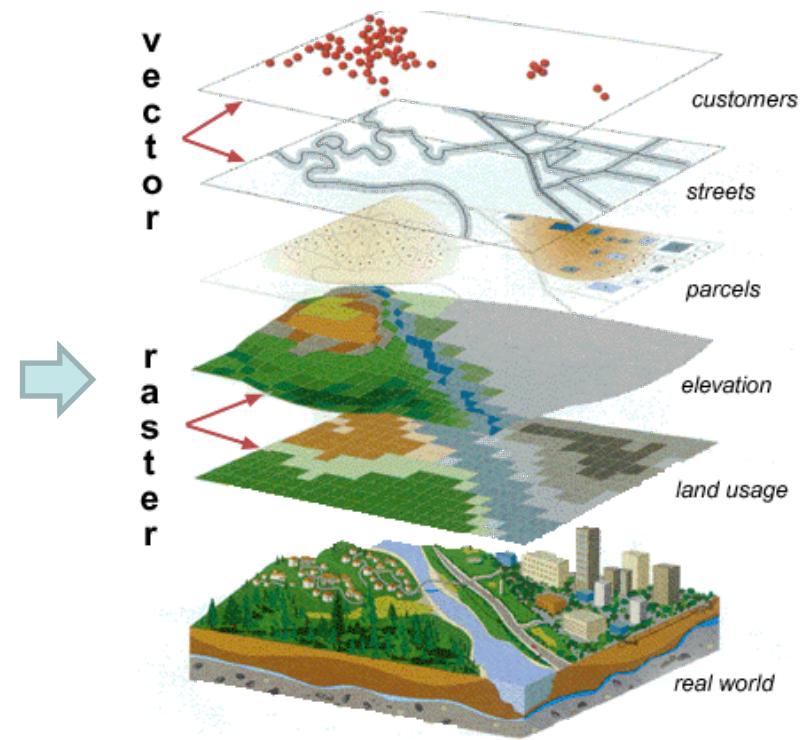


Raster data are good at:

- representing continuous data (e.g., slope, elevation)
- representing multiple feature types (e.g., points, lines, and polygons) as single feature types (cells) – ‘rasterized’ data
- rapid computations (“map algebra”) in which raster layers are treated as elements in mathematical expressions
- suits analysis of multi-layer or multivariate data (e.g., satellite image processing and analysis)
- requires significant disk space

GIS (Geographic Information System) is a software tool for handling spatial data

The raster view of the world	Happy Valley spatial entities	The vector view of the world
	 x x Points: hotels	
	 Lines: ski lifts	
	 Areas: forest	



Space-time context for knowledge discovery

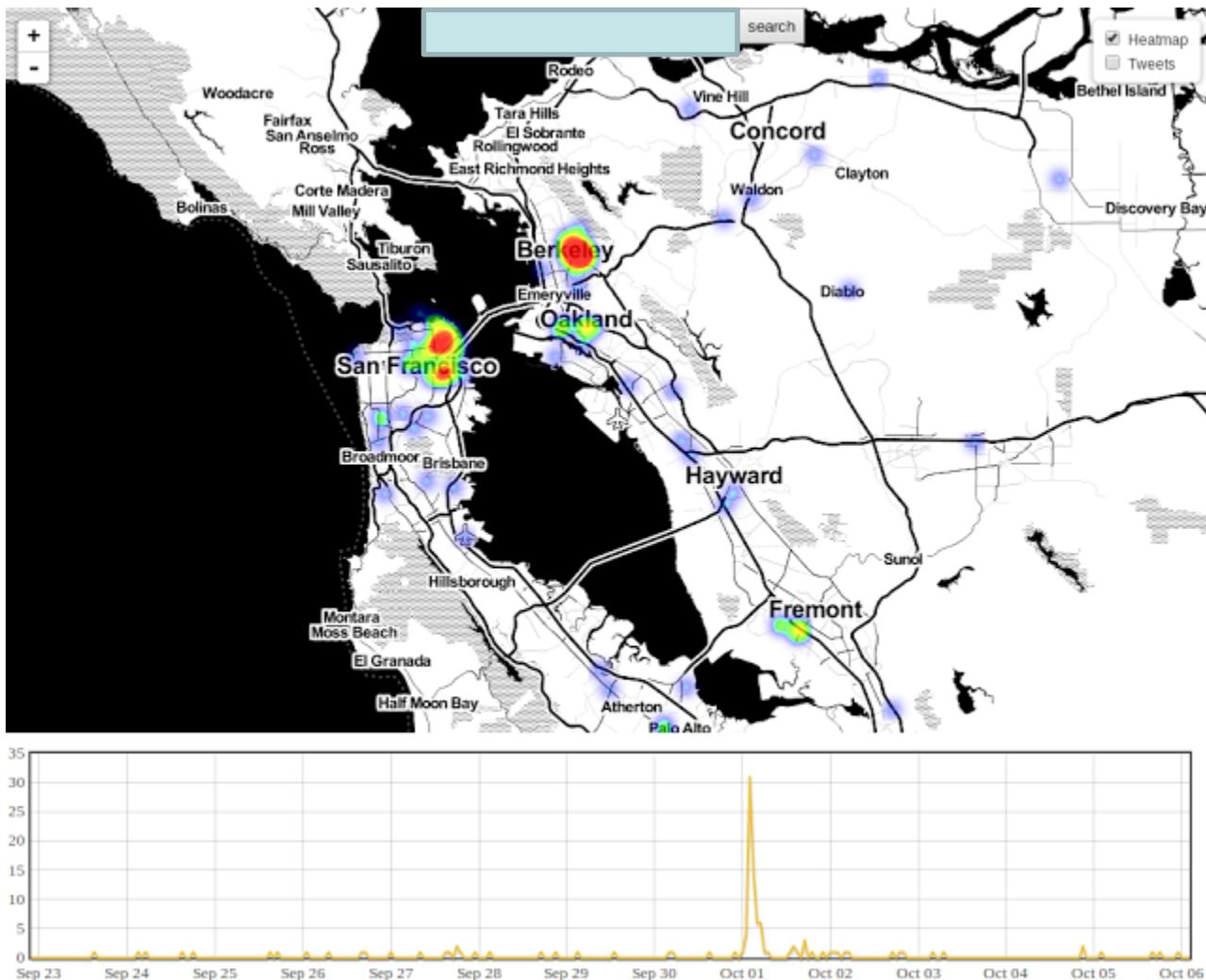


Example 1

Events on Twitter



Space-time context: events on twitter





Space-time context: events on twitter

“Explosion”

San Francisco CP @SanFranciscoCP
At Least 1 Hospitalized After UC Berkeley Power Outage, Chemical Spill & Explosion: BERKELEY (KCBS/KPIX 5) — A...
<http://t.co/r09sVWtCNx>

 Mike M @Michael_N_Cali
WTH Explosion at CAL Berkeley

 serk1 @serk1
Some kind of explosion on central campus at UC Berkeley @berkeleyside

 TheRealDyl @seasickdyl
wow UC Berkeley closed due to huge explosion on campus

 Ashley Cortez @ ashlover
Explosion on Berkley campus!!!!!!!
@@@ http://t.co/azn0fvnjGj

 Roy Shadmon @Roytheboy18
Wishing the best to the students and faculty at Berkeley who witnessed the gas explosion

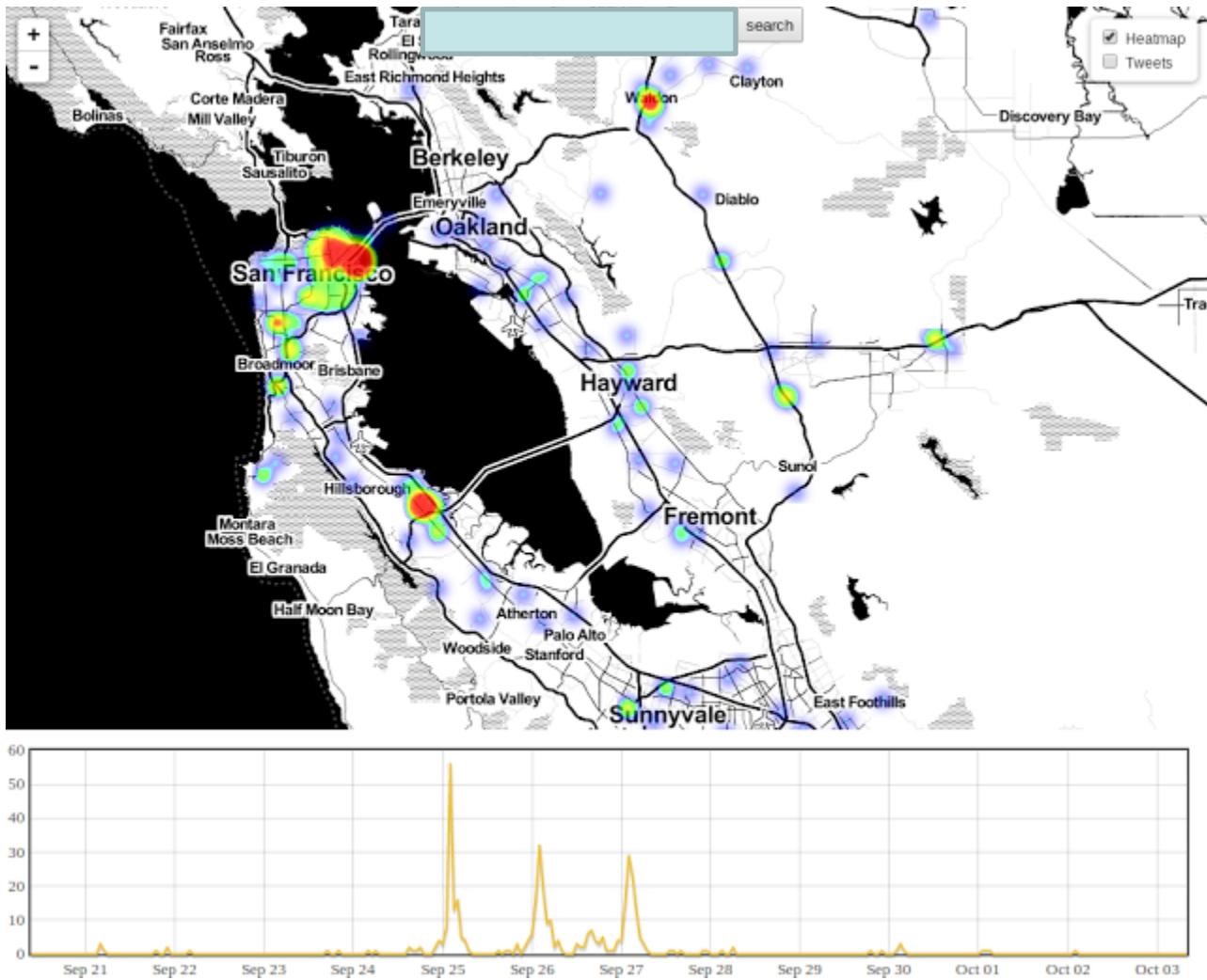
 Carl Onak @CarlOnak
Unconfirmed reports of explosion, police clearing Berkeley campus. I'm at home and safe.

 Jasmeen Bhangu @jasmeenkbhangu
Explosion on Berkeley campus. Officer





Space-time context: events on twitter





Space-time context: events on twitter

“Dodgers VS Giants”

hannah banuelos @HannahJayB
DODGERS VS GIANTS!!! Let's kick
some giant butt dodgers!!! ☺☺☺ #dodgers
#giants #bleedblue...
<http://t.co/EMbCyloxnD>

Savage @Savageman32
@BHOAK510 @mgonz1999
@oakland510 no doubt! Went to a Giants
/Dodgers game and the Giants fans were
WEAK! Dodger fans outnumbered were
louder

San Jose CP @SanJoseCP
Police: Dodgers-Giants rivalry led to
homicide: Jonathan Denver, 24, died of stab
wound after Giants-Dodger game,...
<http://t.co/cVs6dDTXAY>

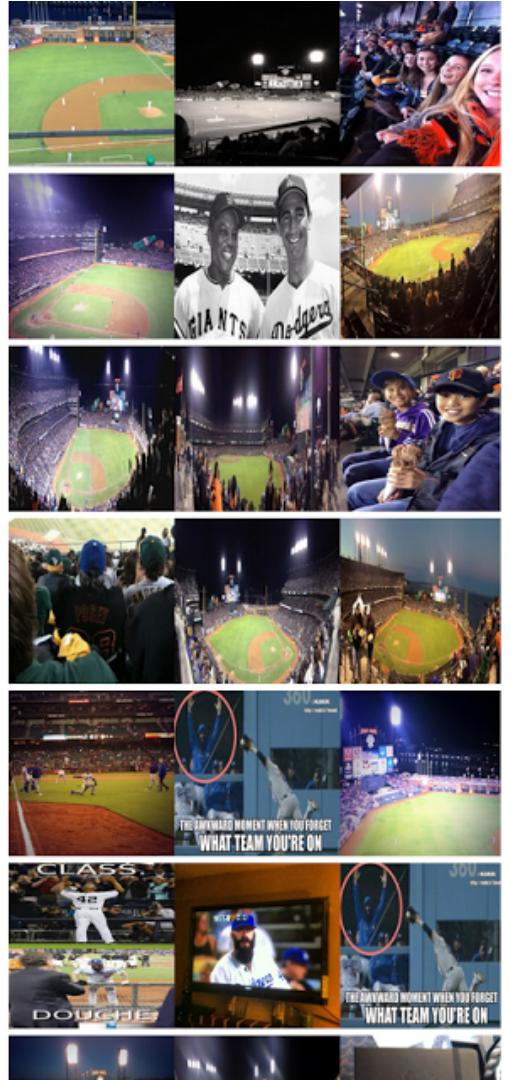
Sara Callahan @sarajcallahan
Go Dodgers! I mean Giants. I meant Giants.
Don't hit me. #sf #giants @ AT&T Park
<http://t.co/U7hr4cyPSN>

Katya canelo @Killuh_Katya
@lovelyjaymora @Dodgers I love me them
Dodgers! Bouta wear my Dodger gear and
get trashed talked by Giants fans lmao it
dont matter tho lol

Heather @khat99
Giants vs. dodgers ! Lets go giants!!!
#mlbbaseball

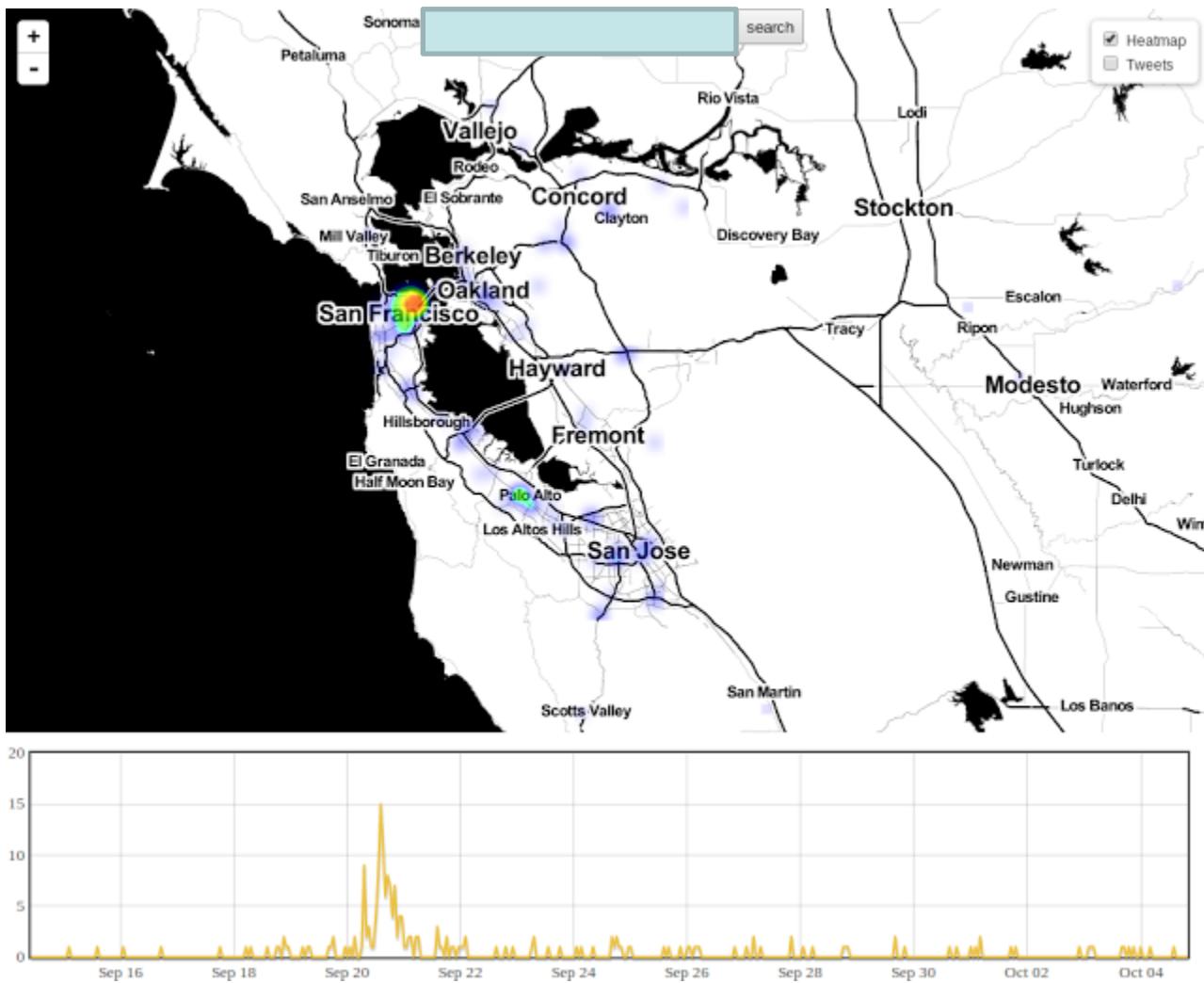
Haleigh Valenta @haleighvalenta
“@Dodgers: #Dodgers defeat the Giants, 2-
1!” ☺

Frederick Lam @fredlam71
At a Giants vs Dodgers game. C'mon
Giants!!! <http://t.co/9695bfIfmqp>





Space-time context: events on twitter





Space-time context: events on twitter

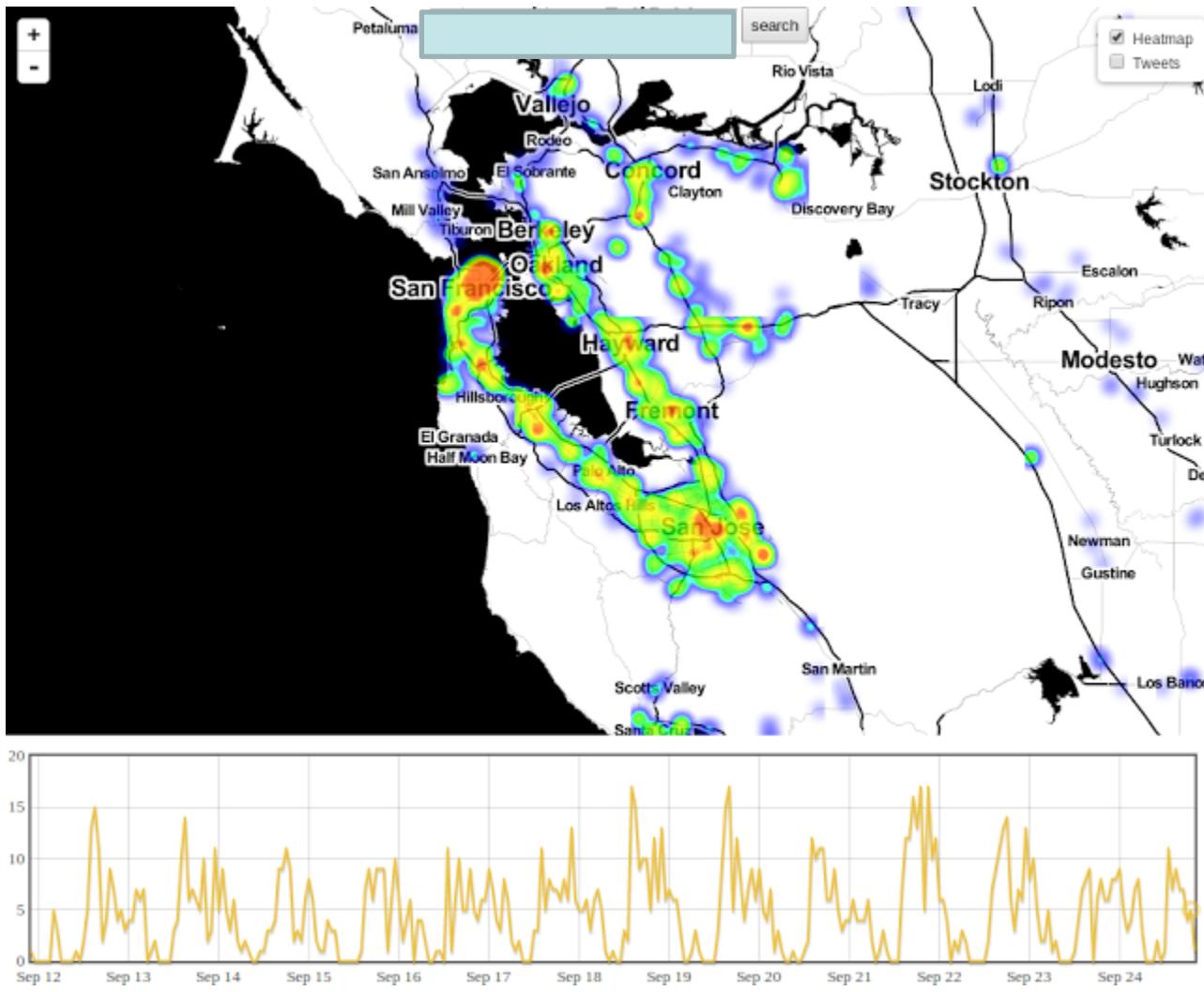
“iPhone5”

- Catherine Hwang P** @findingcat
At the AT&T store here in SF, we r being told all retailers only have black iPhone5S. For silver or gold, go to the Apple store.
#iPhone5S
- Tugce Uhuurgun** @tugceuluurgun
Iphone5s ile Iphone5 arasındaki farkı anlayan el kaldırınsın. Ayrıca renkler de alakasız olmuş bence, yeni bir apple vakası
- Alberto Padilla** @apadillaf
Mi último post: Creando Apps para #ios7 con Corona SDK y adaptando la pantalla al #iphone5 <http://t.co/fTToJ9rGEI> #iPhone5S #app #html5
- Austin** @TheDigitaura
Idiots... #iPhone5S
- COMEIN UP** @MVP_JAYWILL
Gotta get a iPhone5
- Sam Stilwell** @StillSalvatore
#iPhone5S Pre-Ordered!
- OCT4\$ LILKEN** @skinnykvд
I hope she get me the iphone5
- Renée Diaz** @ReneeDiaz28
Need that new #iPhone5S
- mfbecccca** ?? @mfbecccca





Space-time context: events on twitter





Space-time context: events on twitter

“Starbucks”

Ash Chahal. @AshGoesDeep
Starbucks should be closer

NUVIA @nubializethhh
Starbucks ☺

no @cayla_thoughhh
I need Starbucks.

Tristan Morris @tristyyy14
In n out and Starbucks

Princess M †oo @xoxo_monicca
STARBUCKS FINAALLLYYYYY!!!!!!

Daisy Rosales Garcia @DaisyMars
I need Starbucks

Alexis @uh_lex_ees
When does Starbucks open?

ivan ♂ @look_what_daddy
I need Starbucks !!!!!

Myriassa-Victoria @shakarissa
Craving Starbucks

smelly hundy @twitsmawee





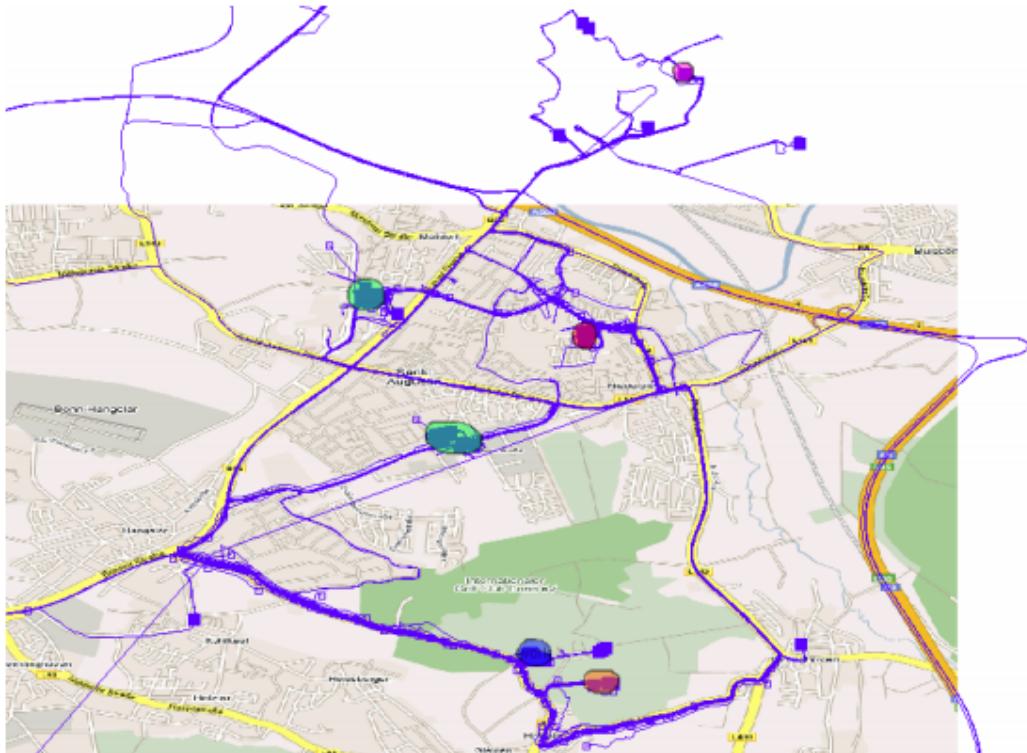
Example 2

GPS tracks

Space-time context for visual analytics

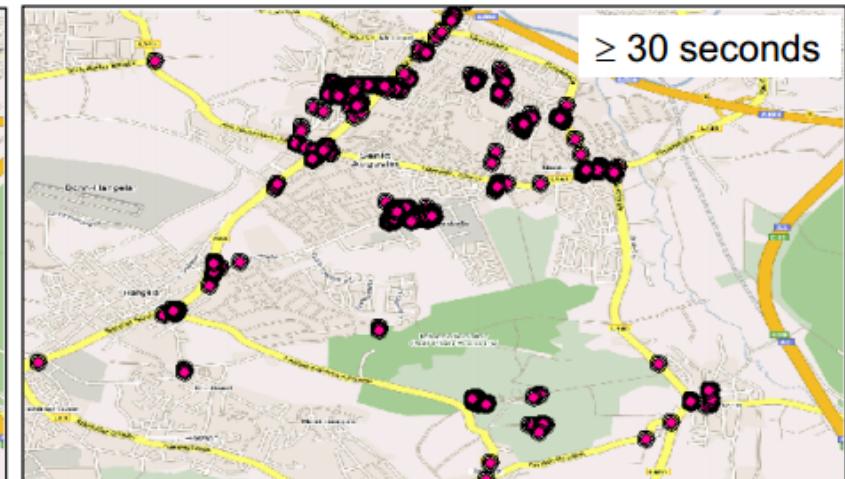
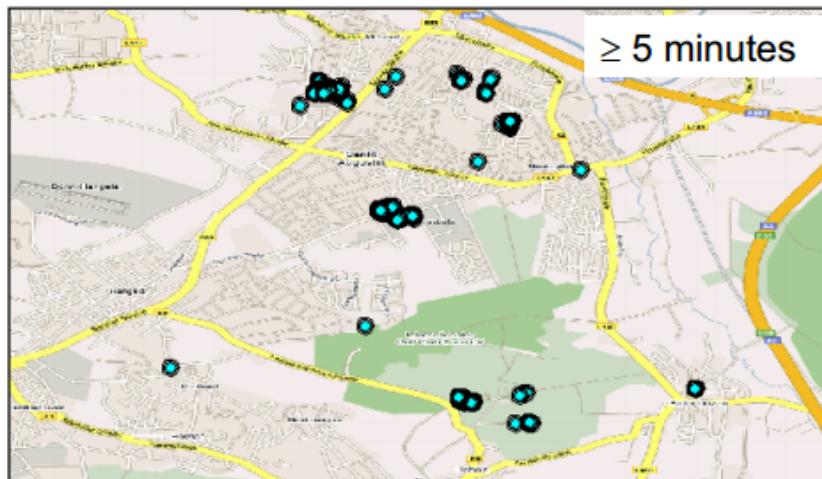
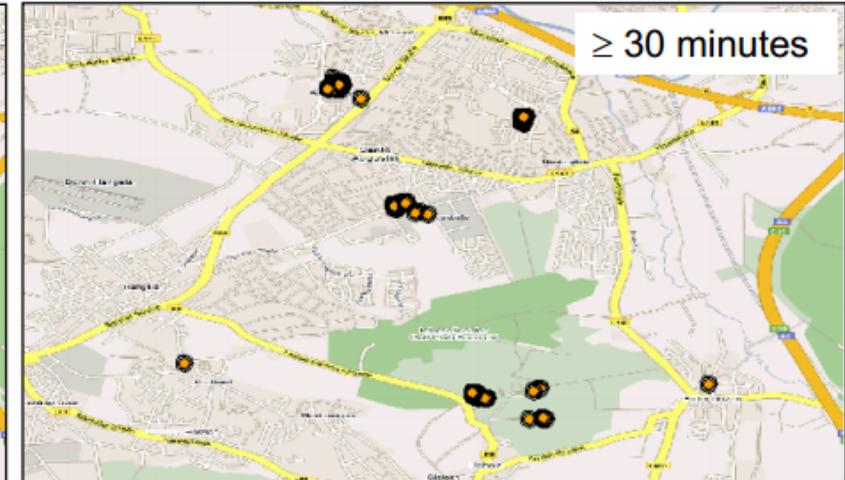
Example of visual analytics: analysis of a personal movement behaviour

- Data: positions of a personal car tracked over a long time period (about a year)
- Task: investigate the movement behaviour of the car owner:
 - Identify significant places and relate them to person's activities
 - Detect and interpret typical trips: sources, destinations, routes, intermediate stops, purposes, ...
 - Detect different routes between the same places; explain when and why each route is chosen
 - ...



Space-time context for visual analytics

Finding significant places: looking at the positions of the stops

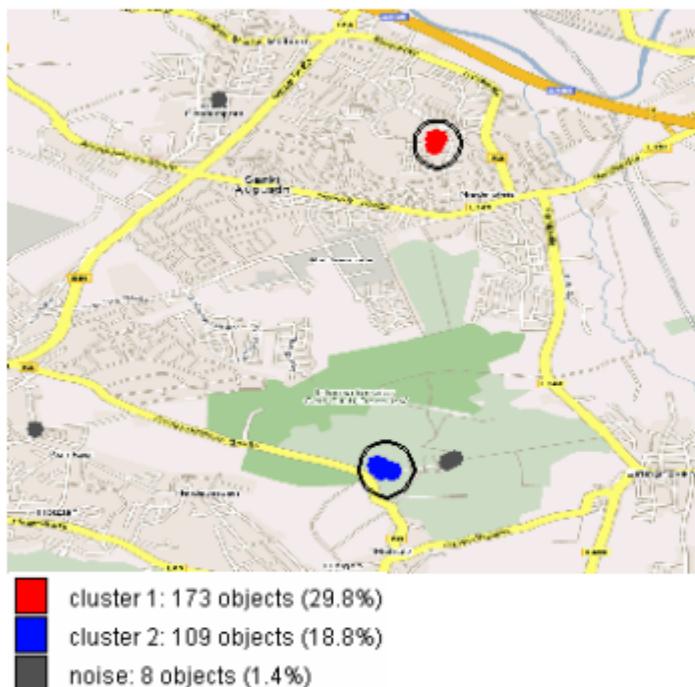


Space-time context for visual analytics

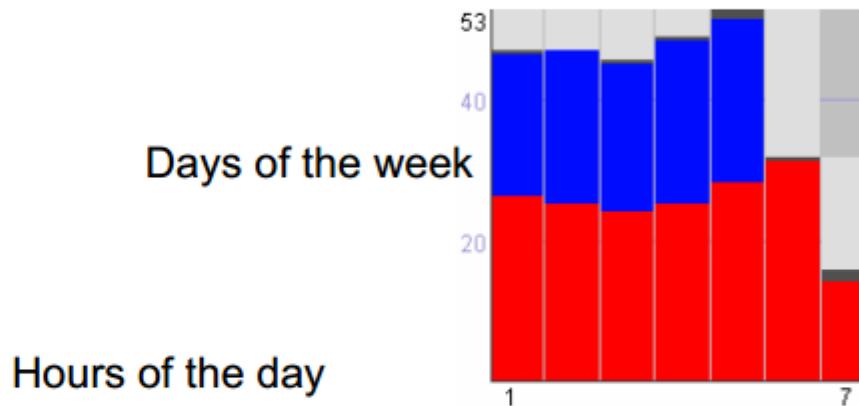
Interpretation of the places of stops

A) Long stops (≥ 3 hours)

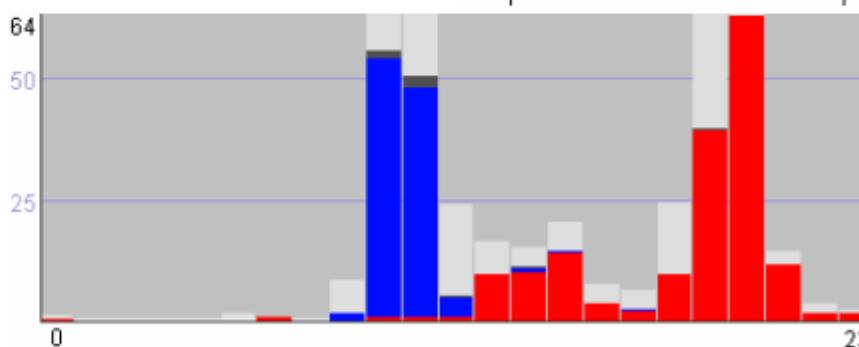
1) Spatial clustering: find repeated stops



2) Look at the days and times of the occurrence



Hours of the day



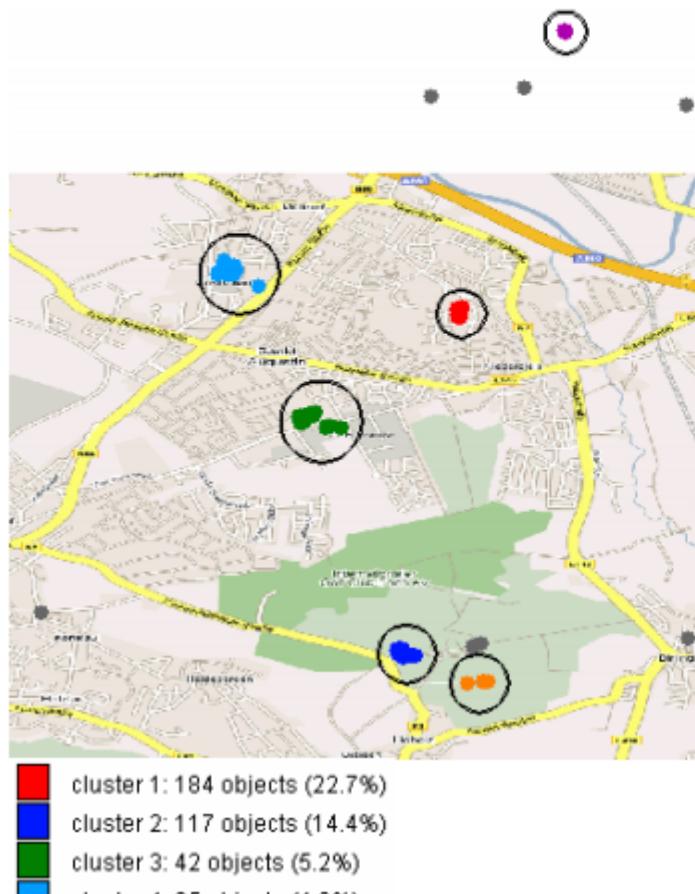
⇒ Red: home, blue: work

Space-time context for visual analytics

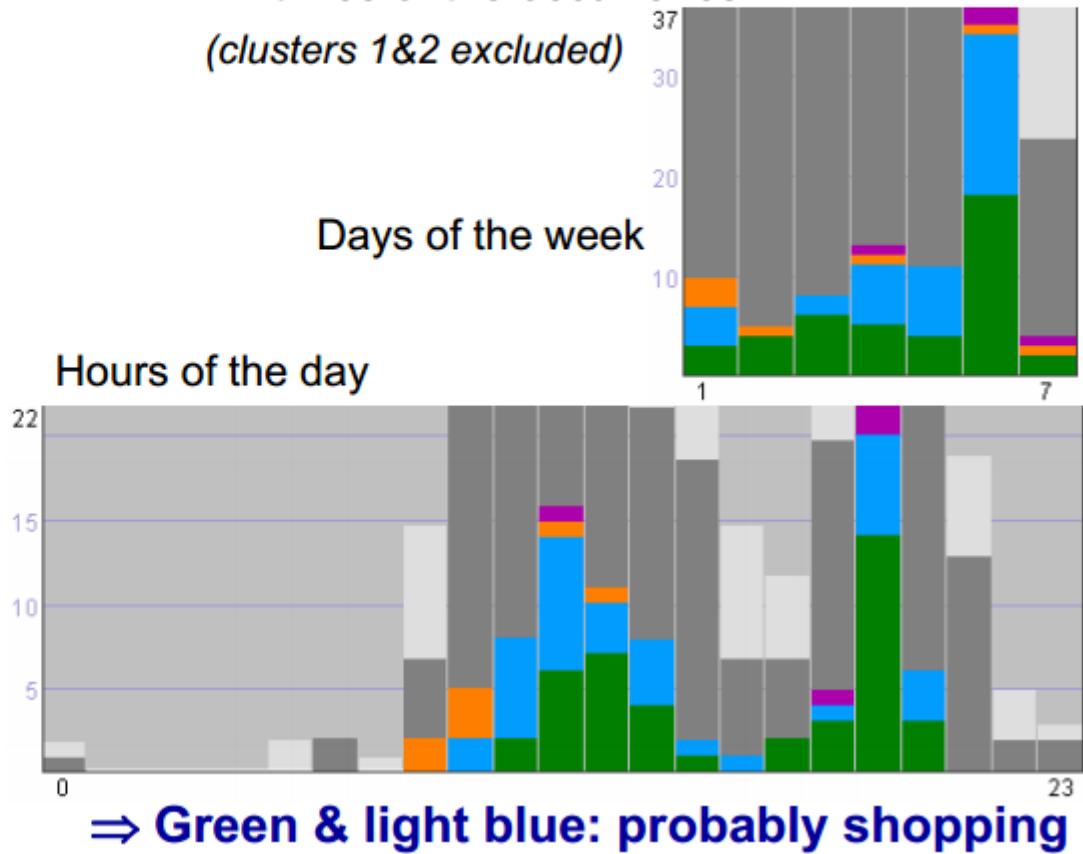
Interpretation of the places of stops

B) Medium stops (≥ 30 minutes)

1) Spatial clustering: find repeated stops



2) Look at the days and times of the occurrence
(clusters 1&2 excluded)

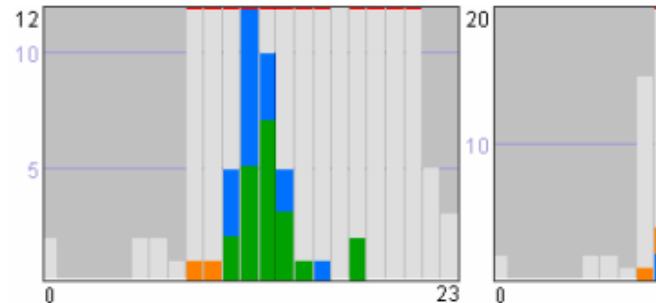
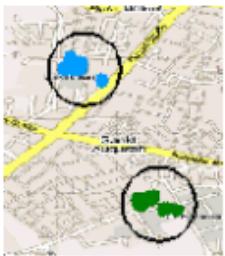




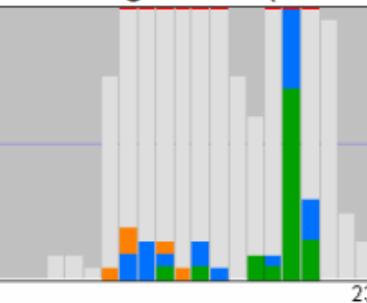
Space-time context for visual analytics

Interpretation of the clusters 3 and 4 (continued)

Saturday and Sunday: the stops mostly occur between 10 and 14



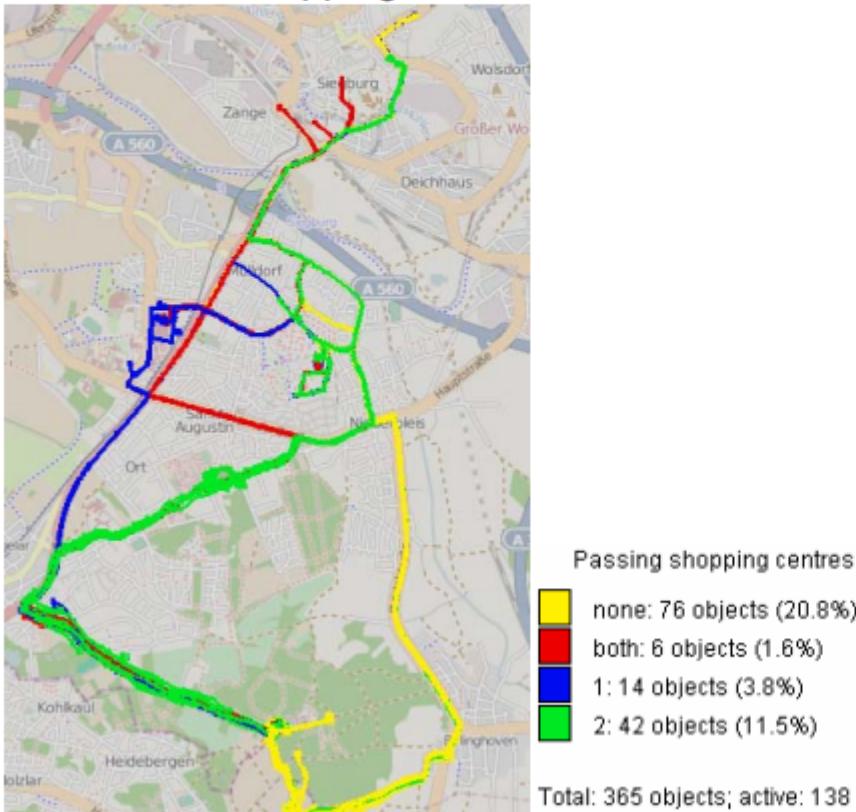
Monday to Friday: the stops mostly occur in the evening hours (max between 18 and 19)



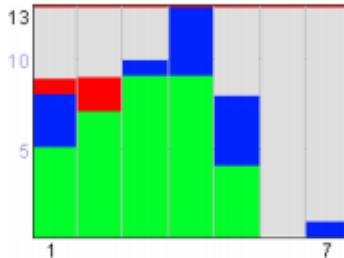
Space-time context for visual analytics

Is there any regularity in preferring one shopping centre to the other?

On the basis of the clustering results, we have interactively classified the trips according to the visits of the shopping centres.

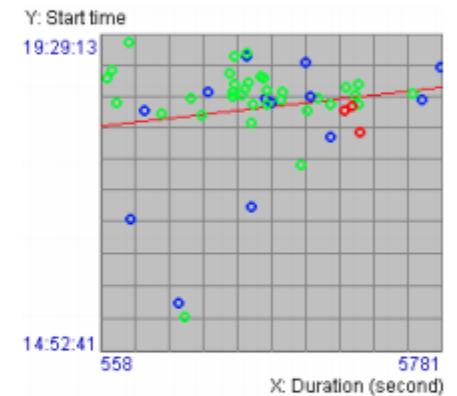


Weekly distribution of the trips



Red: trips through both shopping centres

Start times and durations



- 1) The trips through the shopping centres occur more often on Wednesdays and Thursdays.
- 2) Shopping centre 1 is usually visited on Mondays, Thursdays, and Fridays.
- 3) The starting times and durations of the trips through shopping centres 1 and 2 do not significantly differ.



Space-time context for visual analytics

What could be learned about a car owner using visualization in space-time context:

- The places where the person lives, works, and shops
- The typical routes from home to work, from work to home, to the shopping areas
- The places where the person frequently stops on the way from work to home
- The places where the person may stop on the way from home to work
- The durations of the stops, times spent for visiting the shopping areas
- The times of the trips and of the stops
- How the chosen routes are distributed over the days of the week
- The car owner lives in a small town
- The person has a flexible work schedule
- The person has no small children (concluded from the times of the trips from the work to home)
- The person is often away or sick (judging from the distribution of the trips over the time period, especially in the summer)



Discussion

- Why space and time create patterns and regularities in urban processes?
- How can one observe such regularities in data?
- How can one use such regularities for prediction?
- What is predictability?