

大規模・複雑なデータの 可視化のための集約型表現



筑波大学 システム情報系

三末 和男

misue@cs.tsukuba.ac.jp

目次

- 集約型表現とはどういうものか
- 集約型表現の高度化への取り組み
 - 高次元データ、複合データ、時刻パターン、移動データ
- まとめ
- 公開ソフトウェアの紹介



Aggregation Display

集約型表現とは

パンの重さを可視化したい

□ 毎日のパンの重さを点でプロットする

1か月(30日)分



1年(365日)分

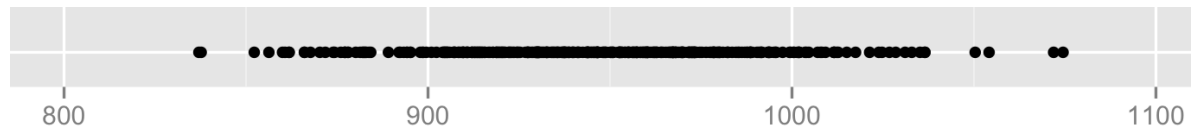


視覚的表現の分類

[Shneiderman, 2008]

□ 項目型 (Atomic)

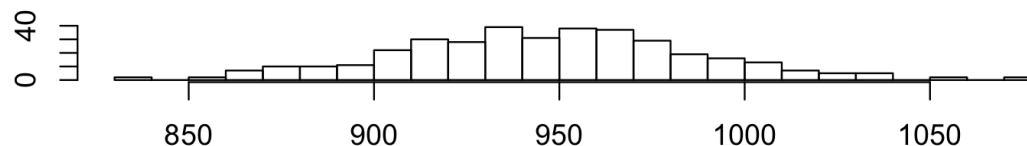
- 各データ項目をひとつの図形で表現



項目数が多くなると視覚的に混雑。ズーム等の併用が必要。

□ 集約型 (Aggregation)

- 複数のデータ項目をひとつの図形で表現



数を視覚属性(色や大きさ)による約束で表現するためスケーラブル。

□ 密度型 (Density)

- データの密度分布を色などで表現



スケーラブル。ただし、基本となる表現が位置に基づく必要がある。

ところで、 そもそも可視化に何を期待するか？

- Overview (全体的ビュー、大局的ビュー)
 - 理解の順序の違い (視覚的表現の特徴)
 - 記述言語: 部分から全体: 文字 → 単語 → 文 → 文章
 - 視覚的表現: 全体から部分: 全体の概形 → 部分 → 詳細
 - マントラが示すもの
 - Overview first, zoom and filter, then details-on-demand.
- リッチなOverview (全体を俯瞰しながらも詳しく)
 - データが複雑化している
 - 単純化された概観ではあまり役に立たない

集約型表現の課題

- 全体を俯瞰しながらも詳しく
 - 基本的な集約型表現は大規模データの量的側面をうまく表現する。
 - もちろん、データの集約を適切に行い、視覚的表現を適切に割り当てる必要がある。
 - しかし、複雑さについては必ずしもうまく対応できない。
- 複雑なデータをいかに表現するか

(ここで言う) 複雑なデータとは

- 多変量
 - 高次元(多変量)
 - 多値カテゴリデータ
- 複合的
 - 階層構造 + 多変量データ
 - 階層構造 + 時系列データ
 - 移動(位置×位置×量×属性)

集約型表現の発展形への取り組み

- 高次元データ: Colored Mosaic Matrix
- 複合データ: SaaG: Series at a Glance
- 時刻パターン: ChronoView
- 移動データ: Amoeba

多高次元データのための集約型表現

Colored Mosaic Matrix

多次元データ

- 複数の次元(変量)を持つデータ。

- (例) 気象データ

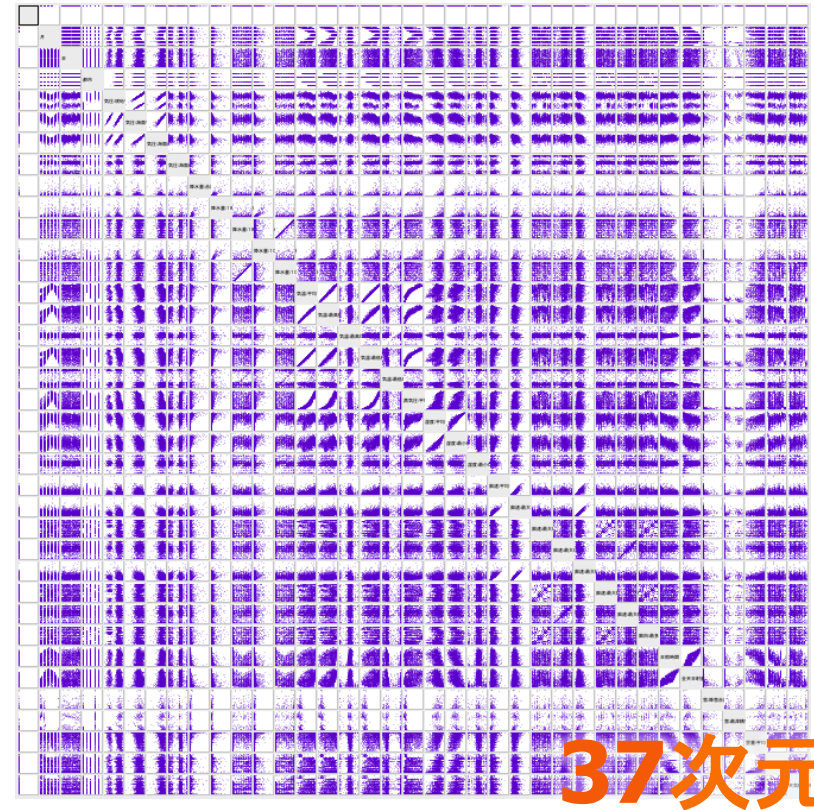
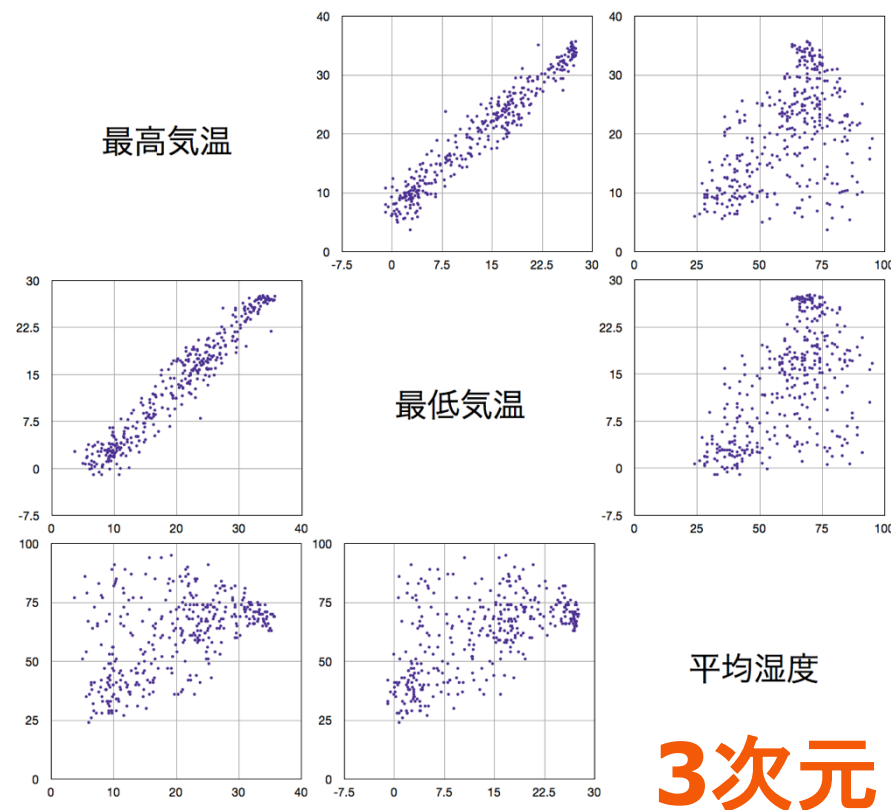
日付	観測地	最高気温	最低気温	降水量	湿度
2013/01/13	つくば市	13.8℃	-4.9℃	0mm	62%
2013/01/14	つくば市	7.8℃	-0.2℃	41mm	89%
2013/01/15	つくば市	6.3℃	-2.4℃	0mm	67%

...

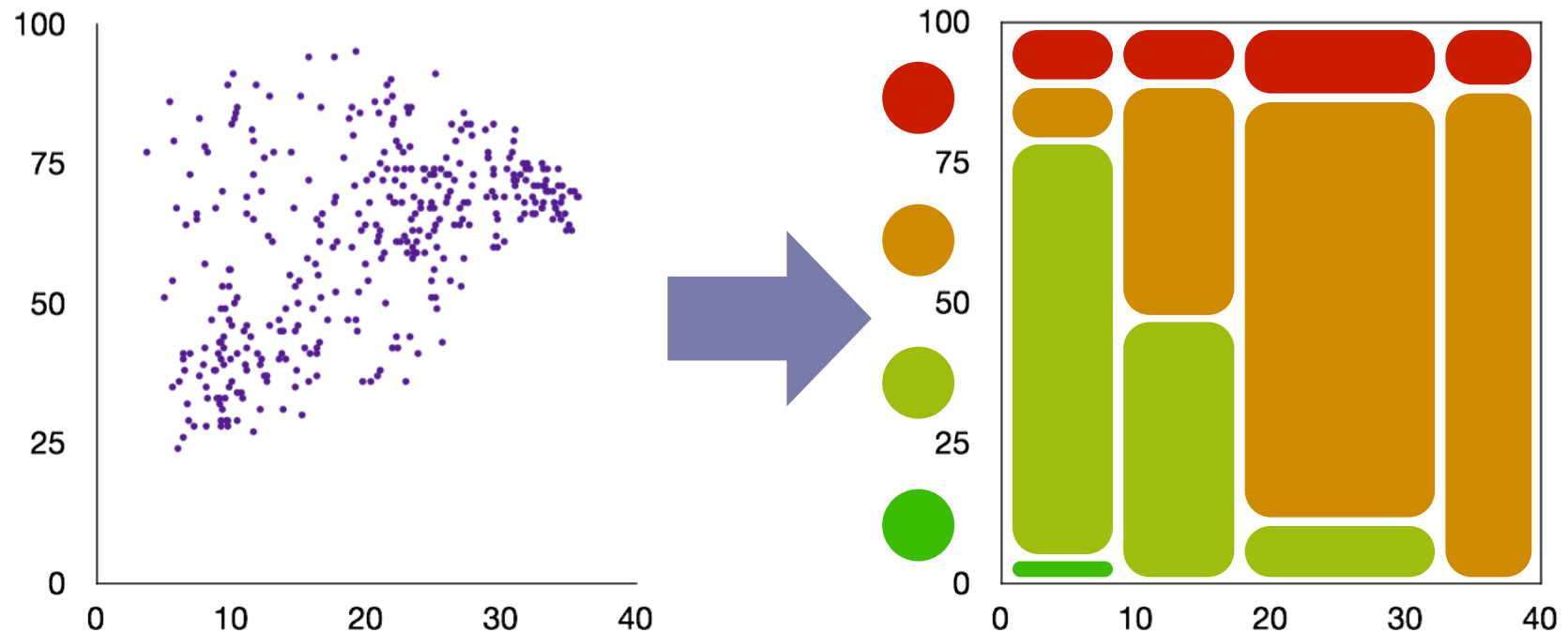
- 各種センサーデータ、ログデータなど、多次元データ(多変量データ)は数多く存在する。
- 特に30次元を越えるデータを扱うことを考える。

散布図行列による可視化

- 散布図行列: 次元の組毎に描いた散布図を行列状に配置。多次元データの関係を一覧できる。

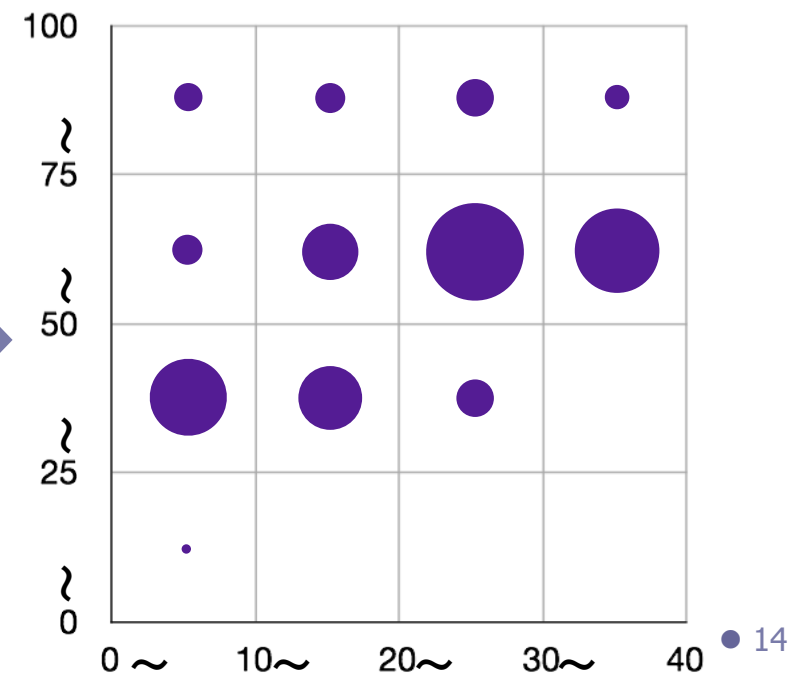
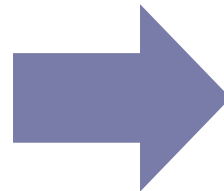
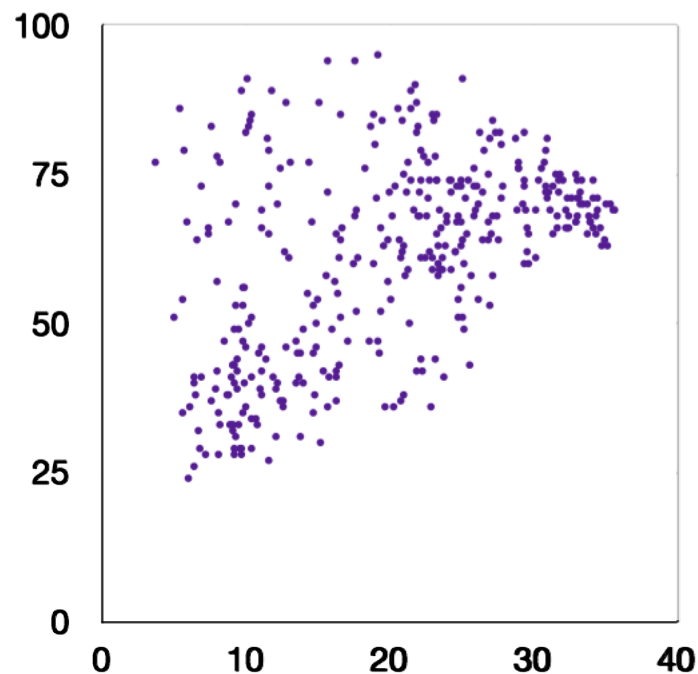


散布図を集約型表現に



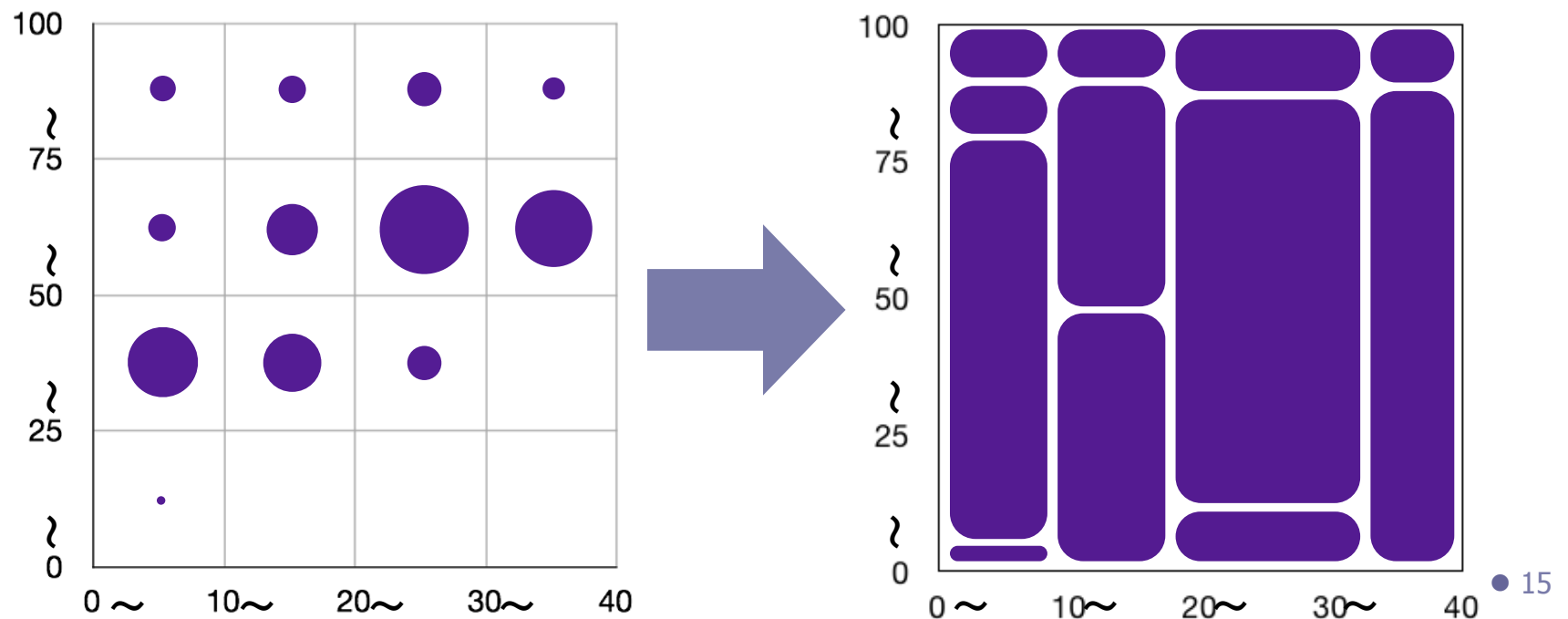
量的データのカテゴリ化

- 量的データの次元については、値域を細分化した上で、各値域をカテゴリとして扱う
- 少量のマーカー(点)でデータを表現できる



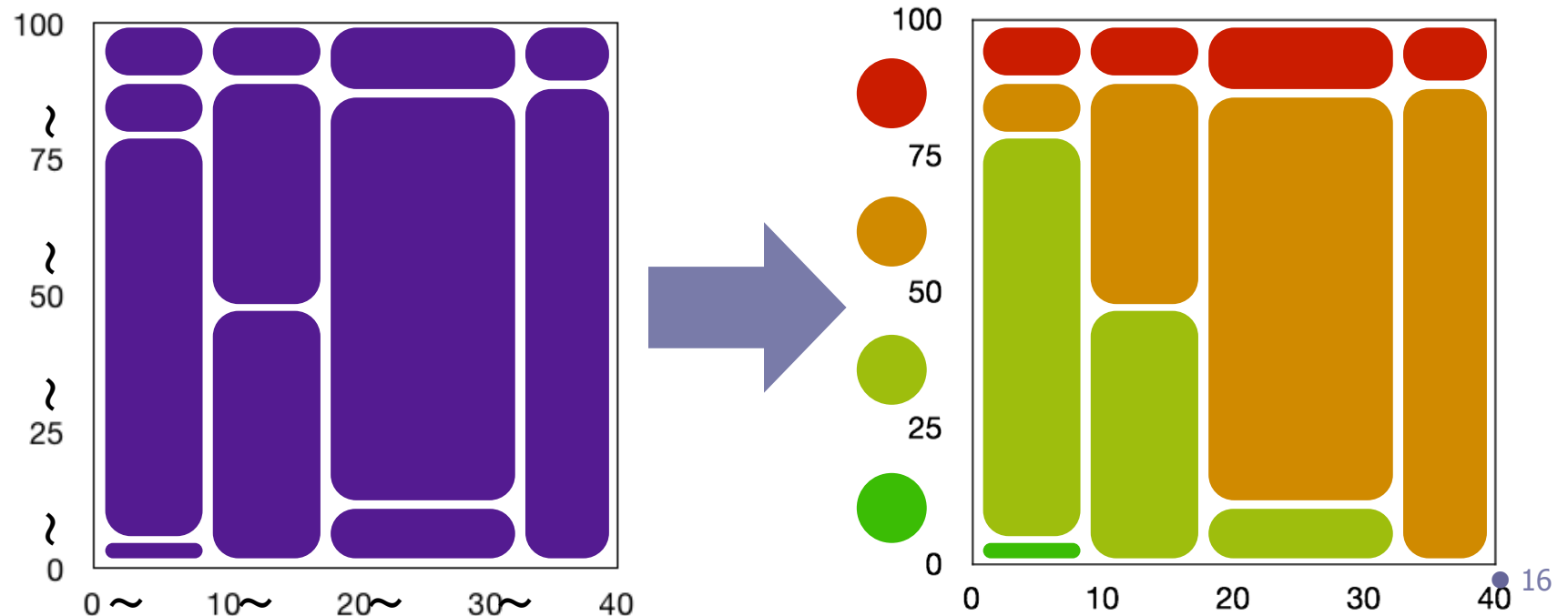
Mosaic Plot [Friendly, 1994]

- 空間充填型のカテゴリデータ可視化手法
- カテゴリの比率に応じて領域を矩形に分割する
 - 2次元データの分布を各矩形の面積で表現する



色付きMosaic Plot

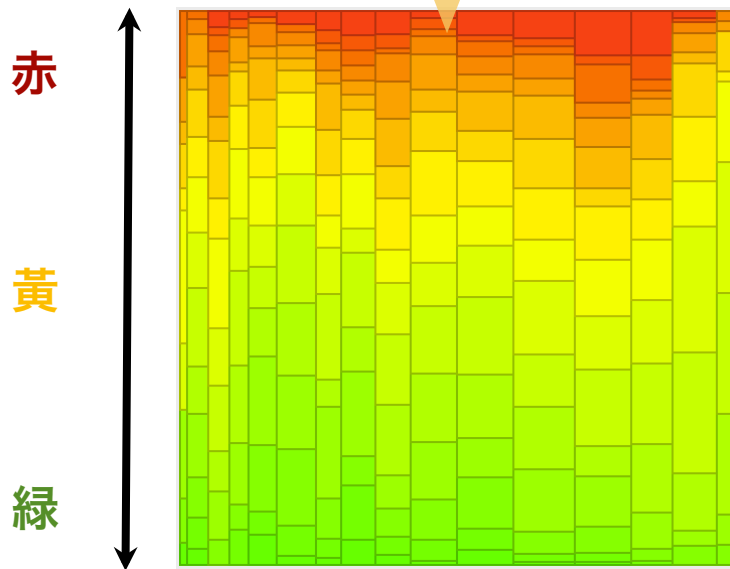
- 小領域での識別を容易にするための色付けを行う
- 2データの分布や特徴を、色彩の模様から読み取る



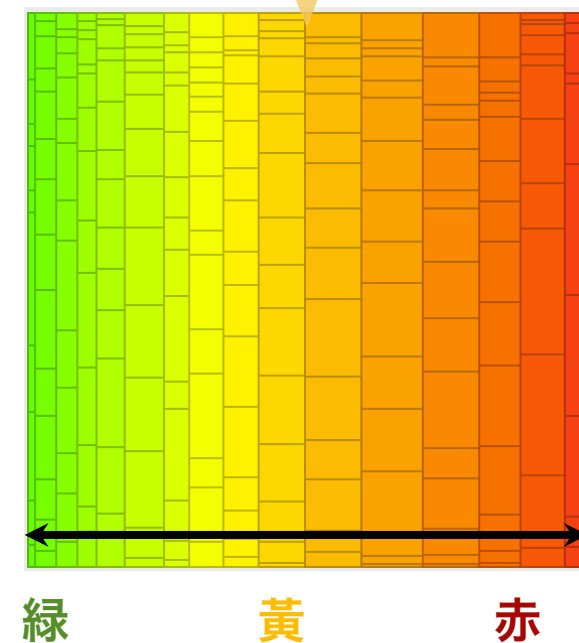
カテゴリの分布に着目した色付け

- 矩形に対応するカテゴリを元に色を決定する
 - 各カテゴリの割合からデータ分布を把握できる

Y軸カテゴリに着目した例



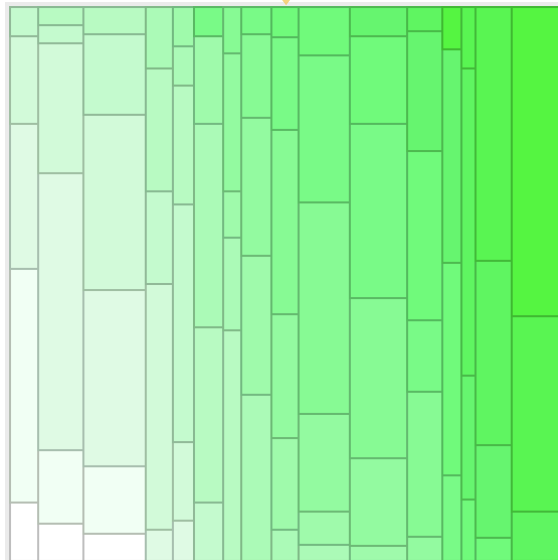
X軸カテゴリに着目した例



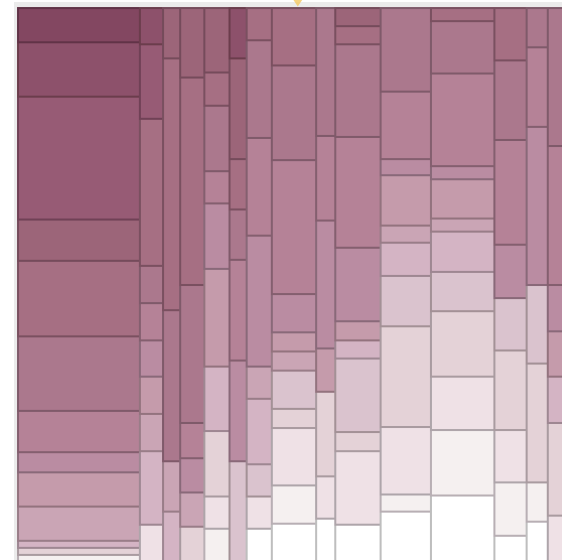
次元間の相関に着目した色付け

- 相関係数を利用して色を決定する
 - 色相: 相関の正負
 - 明度と彩度: 相関の強弱とY軸カテゴリの区別

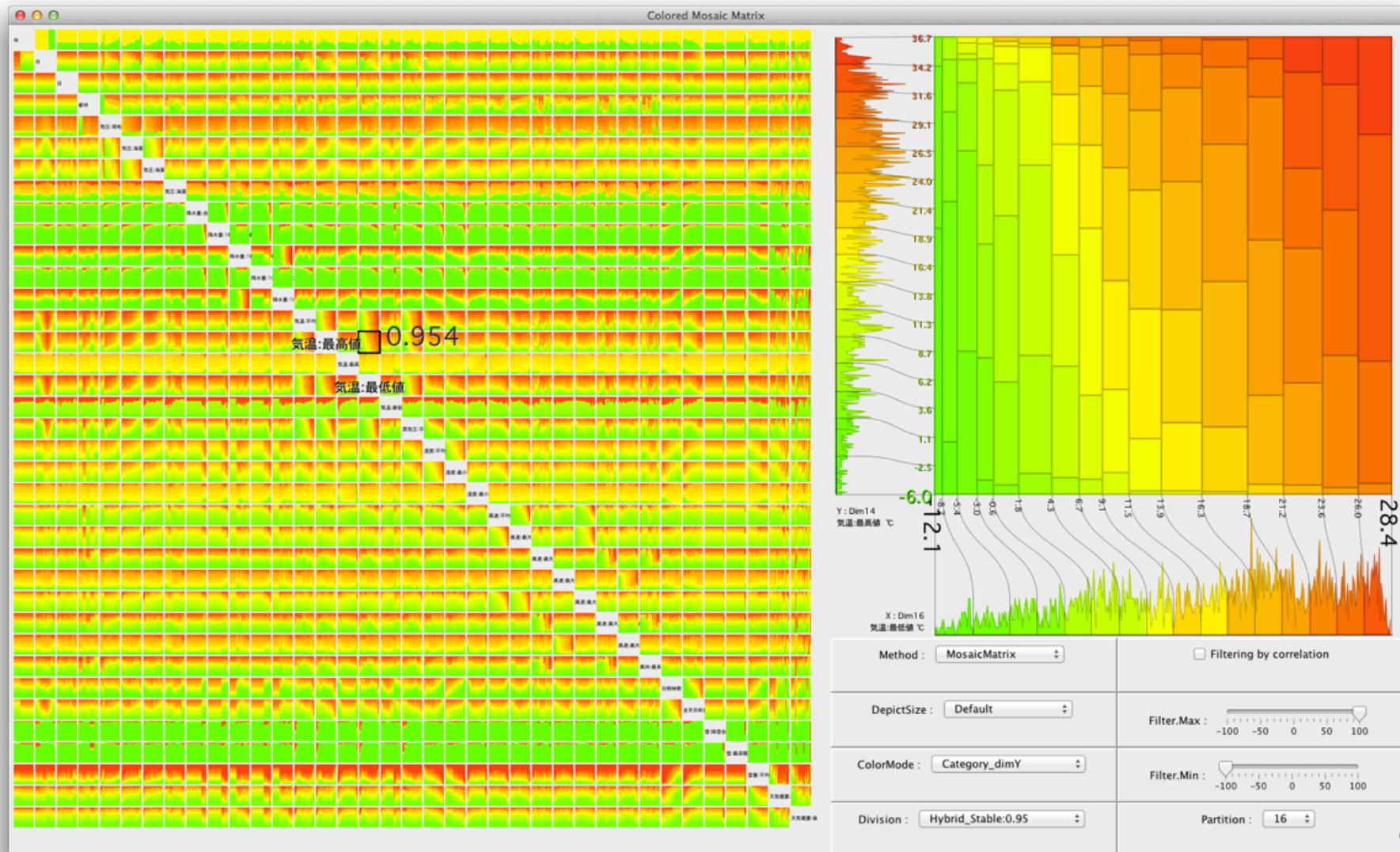
正の相関の例



負の相関の例



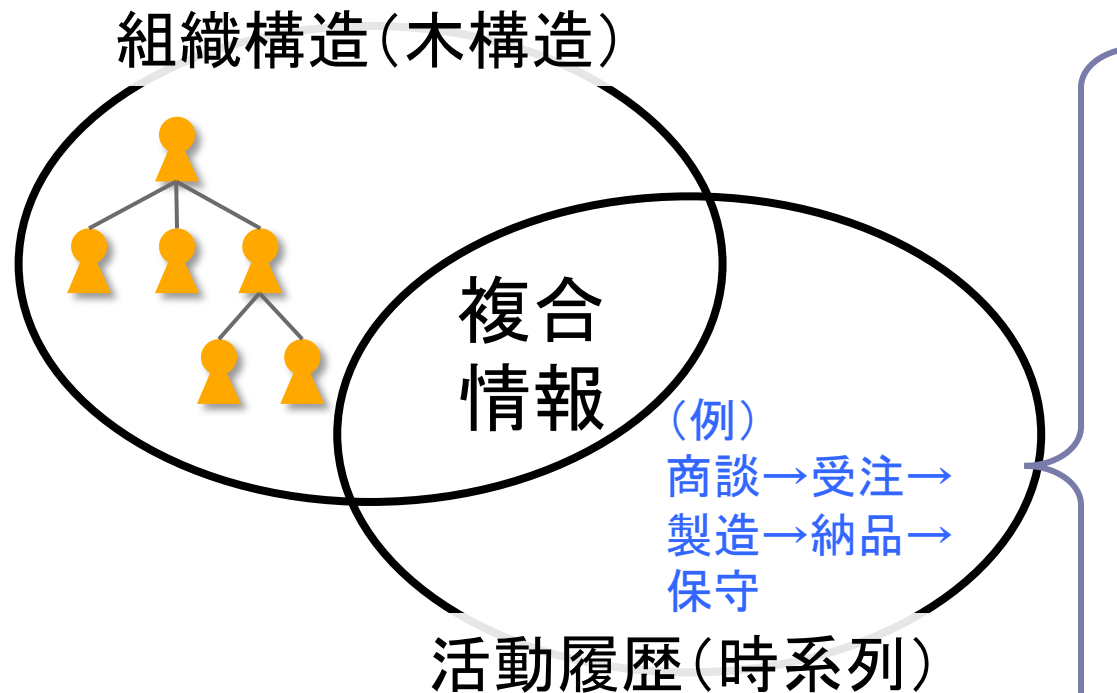
ツールの開発



大規模組織の活動の可視化

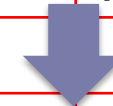
SaaG: Series at a Glance

組織における複合情報

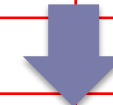


チケット(活動履歴)

Update	`08/04/09
Status	New
Assigned to	(None)



Update	`08/04/10
Status	Assigned
Assigned to	Yazaki



Update	`08/04/17
Status	Closed
Assigned to	Yazaki

属性

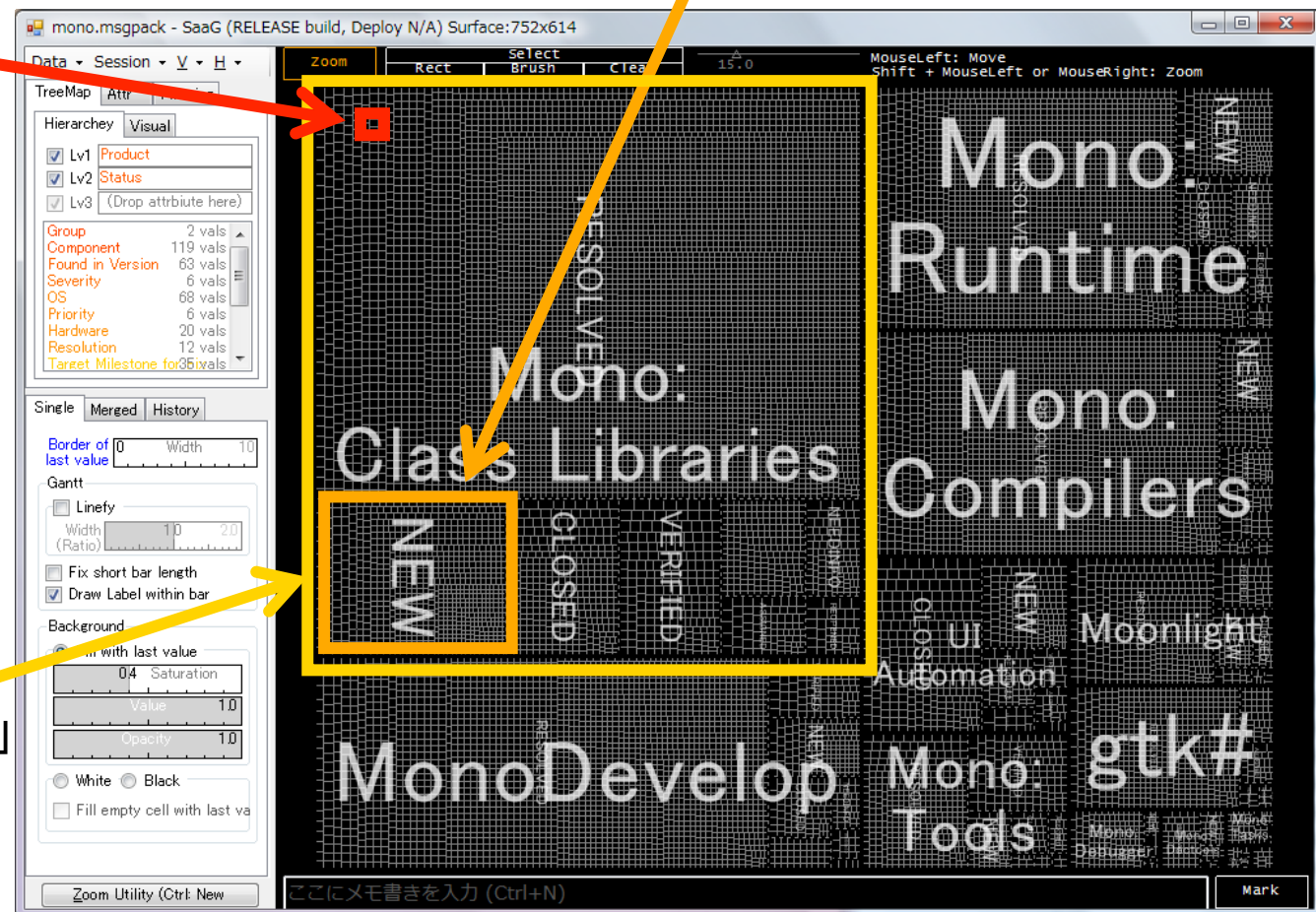
属性値 • 21

組織構造のTreemapによる可視化

製品「Class Library」に関する活動
の中で、新規(NEW)状態のもの

ひとつの矩形がひとつ活動(チケット)を表現。

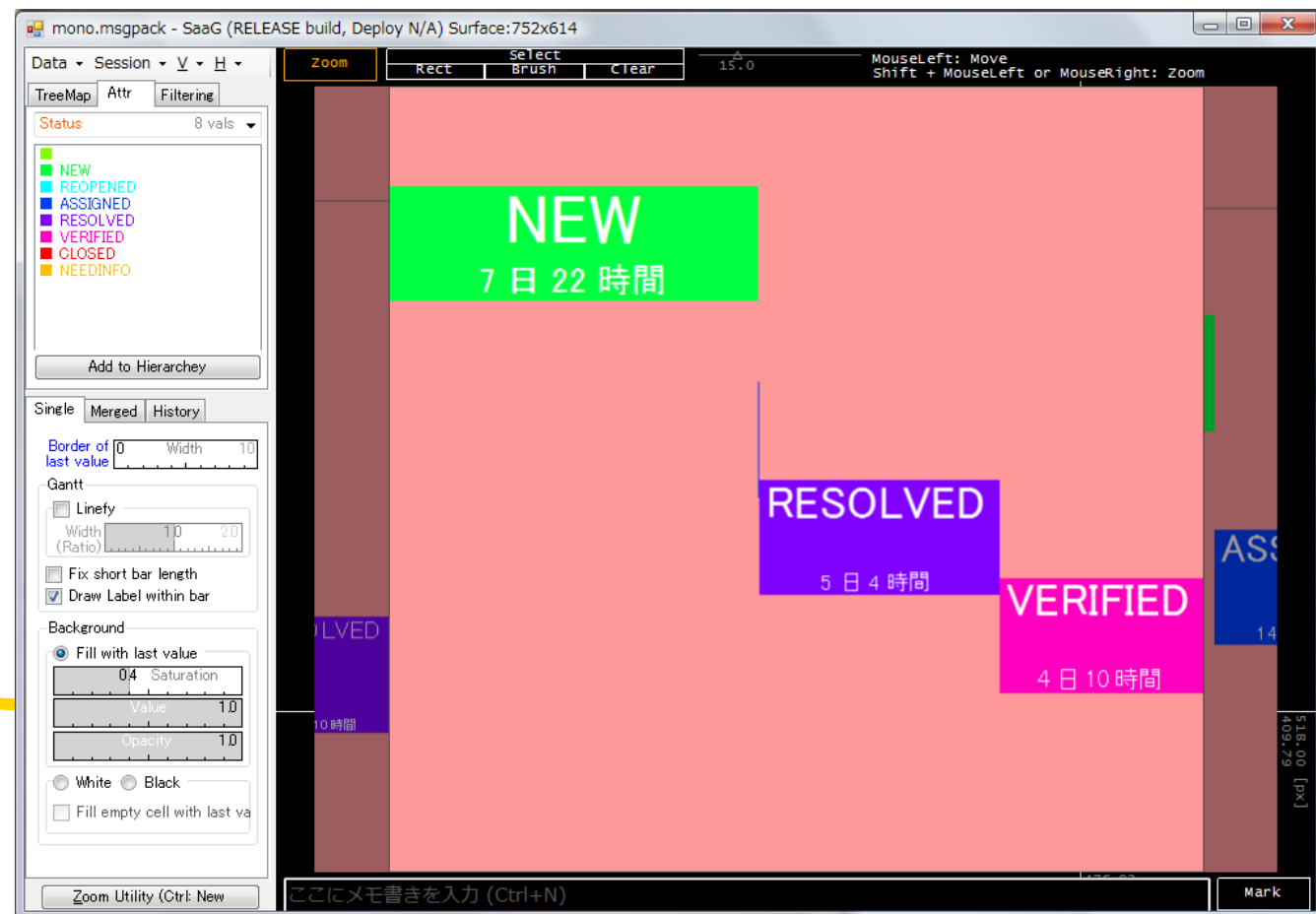
製品「Class Library」に関する活動



活動履歴の埋め込み

ガントチャートにより活動履歴を見ることができる。

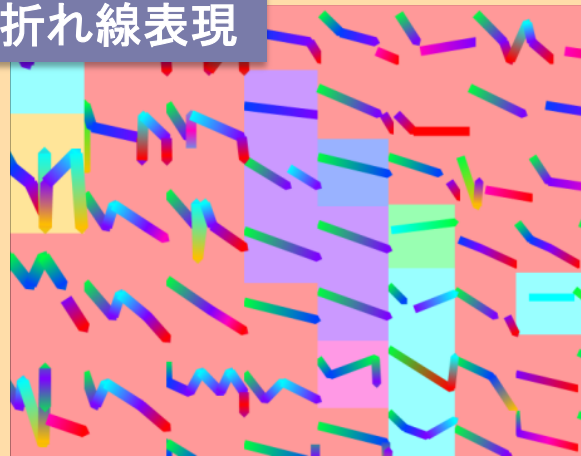
製品「Mono Tools」
を拡大すると



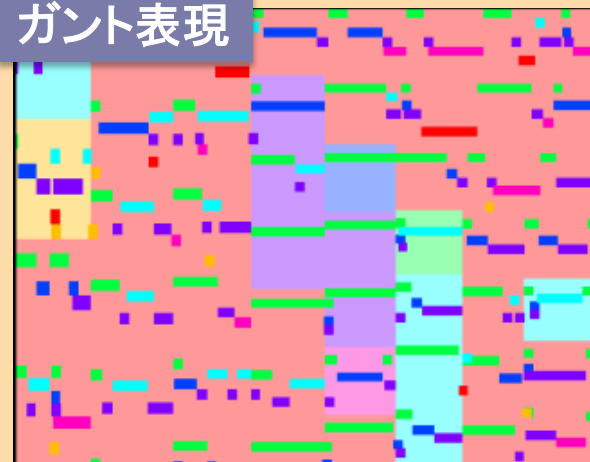
活動履歴の表現

個々のチケット
を個別の矩形
内で表現

折れ線表現

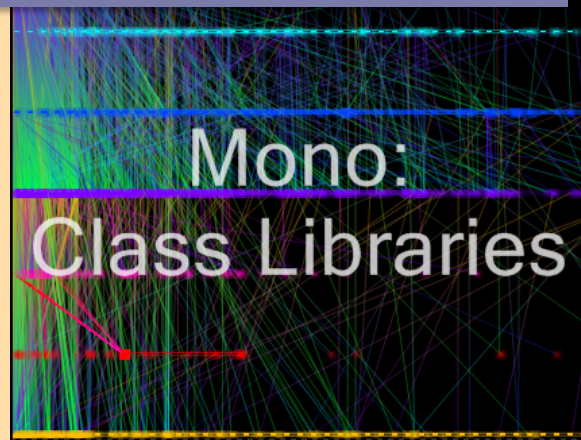


ガント表現

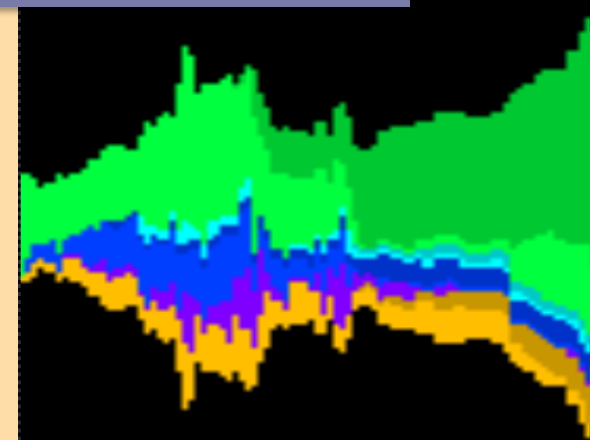


チケット群をひ
とつの矩形内で
表現

折れ線表現＋重ね合わせ

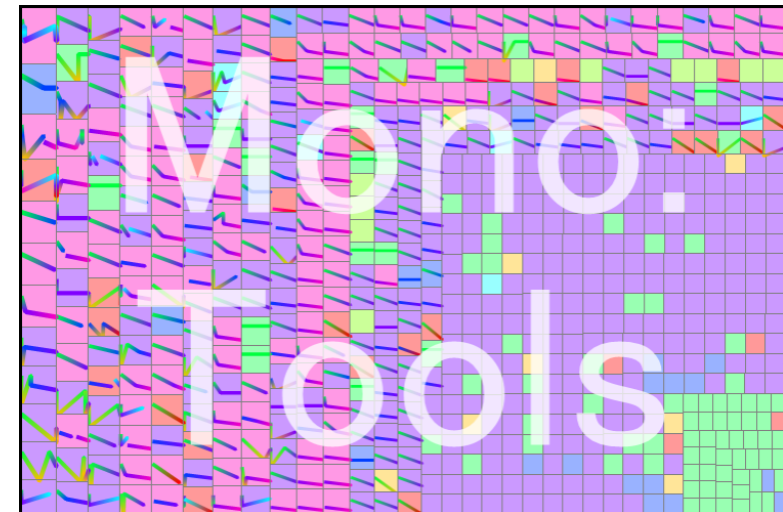
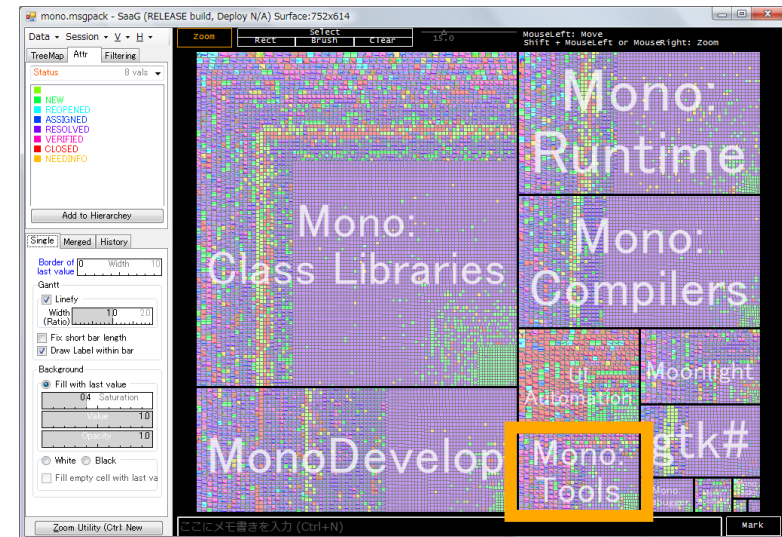


リバー(河川)表現



ガント表現／折れ線表現の並び

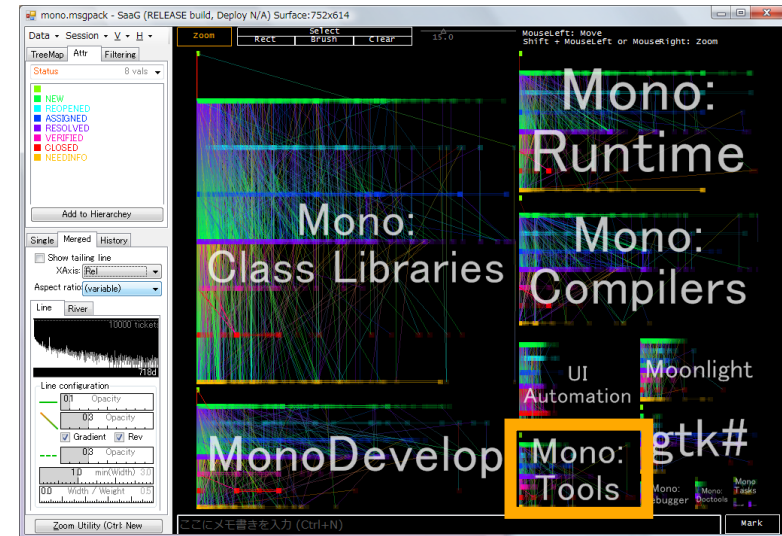
- 折れ線表現をタイル状に並べることでチケット群を表現
- ズームインすれば、チャート毎の形や色のパタンにより傾向を読み取れる。
- ズームアウトした場合には、背景色により現在の状態の分布を読み取れる。



折れ線表現の重ね合わせ

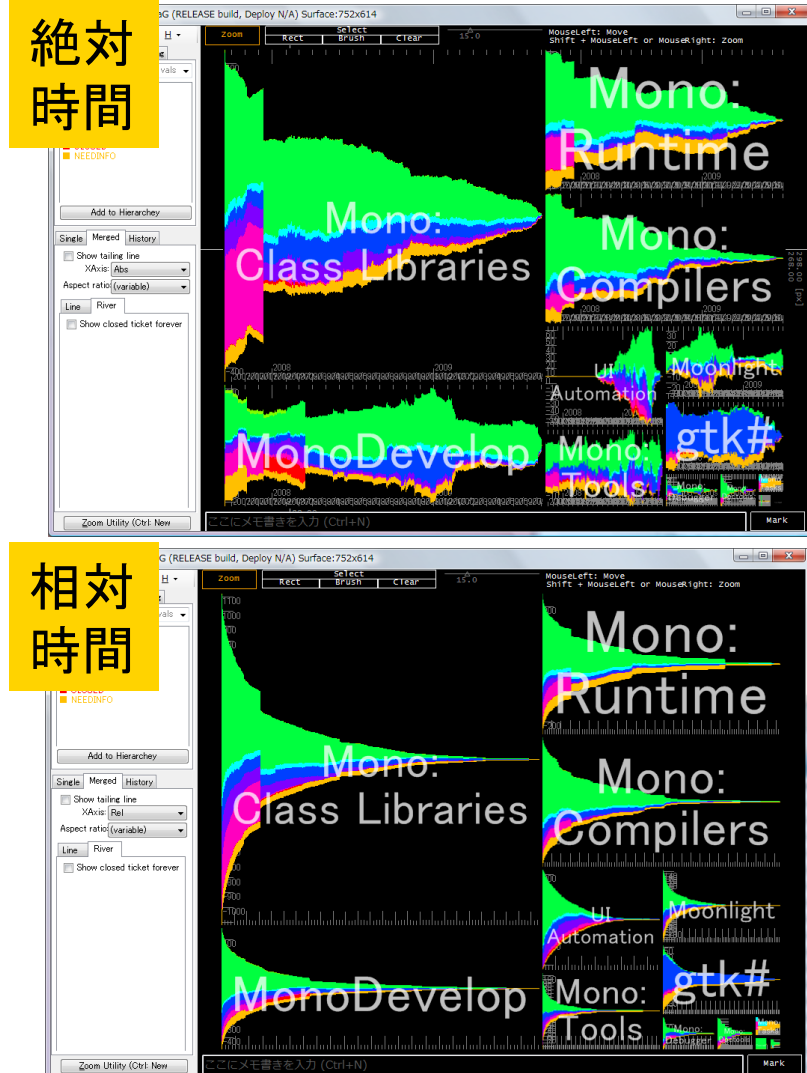
- 折れ線表現を同じ時間軸に重ね合わせてチケット群を表現
- 線の密度からおおまかな傾向や例外が読み取れる。
- 元や行き先がわかるように逆のグラデーションを採用

NEW
RESOLVED



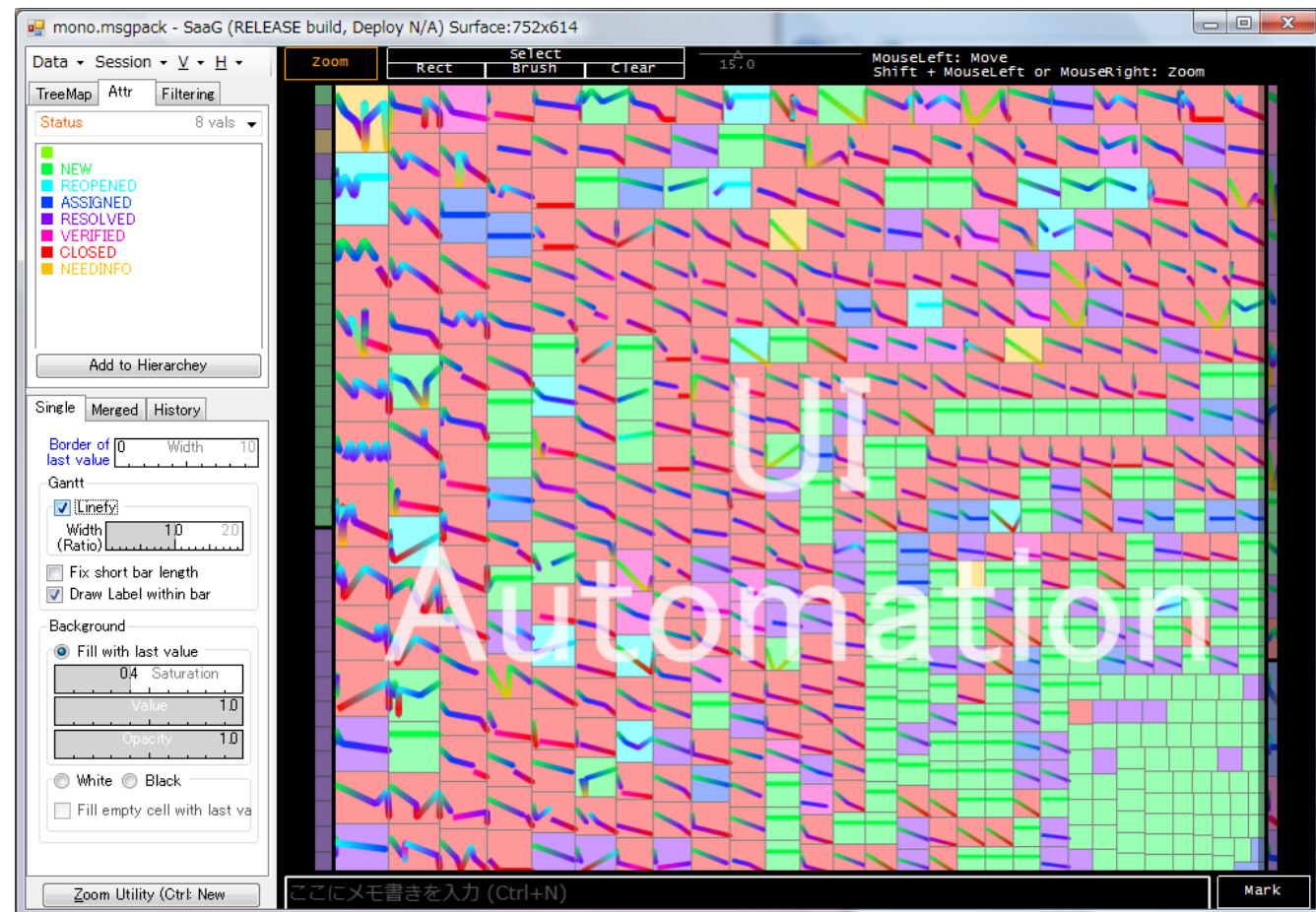
リバー表現

- チケットの量的変化を積み上げ型チャートで表現。
- ある時点のチケットの量、量の変化を明確に捉えることができる。
- 時間軸は絶対時間と開始からの相対時間を切り替え可能。
 - 折れ線表現でも可能



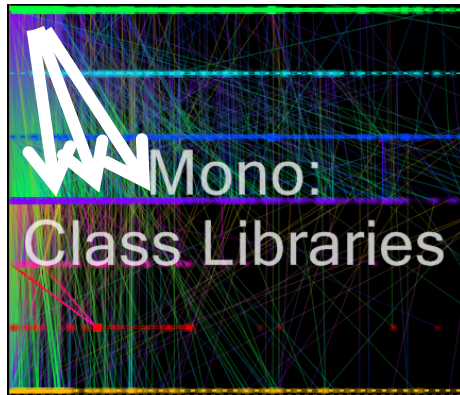
(例) 背景色で見える特徴

- UI Automationは他の製品とは色合いが異なる

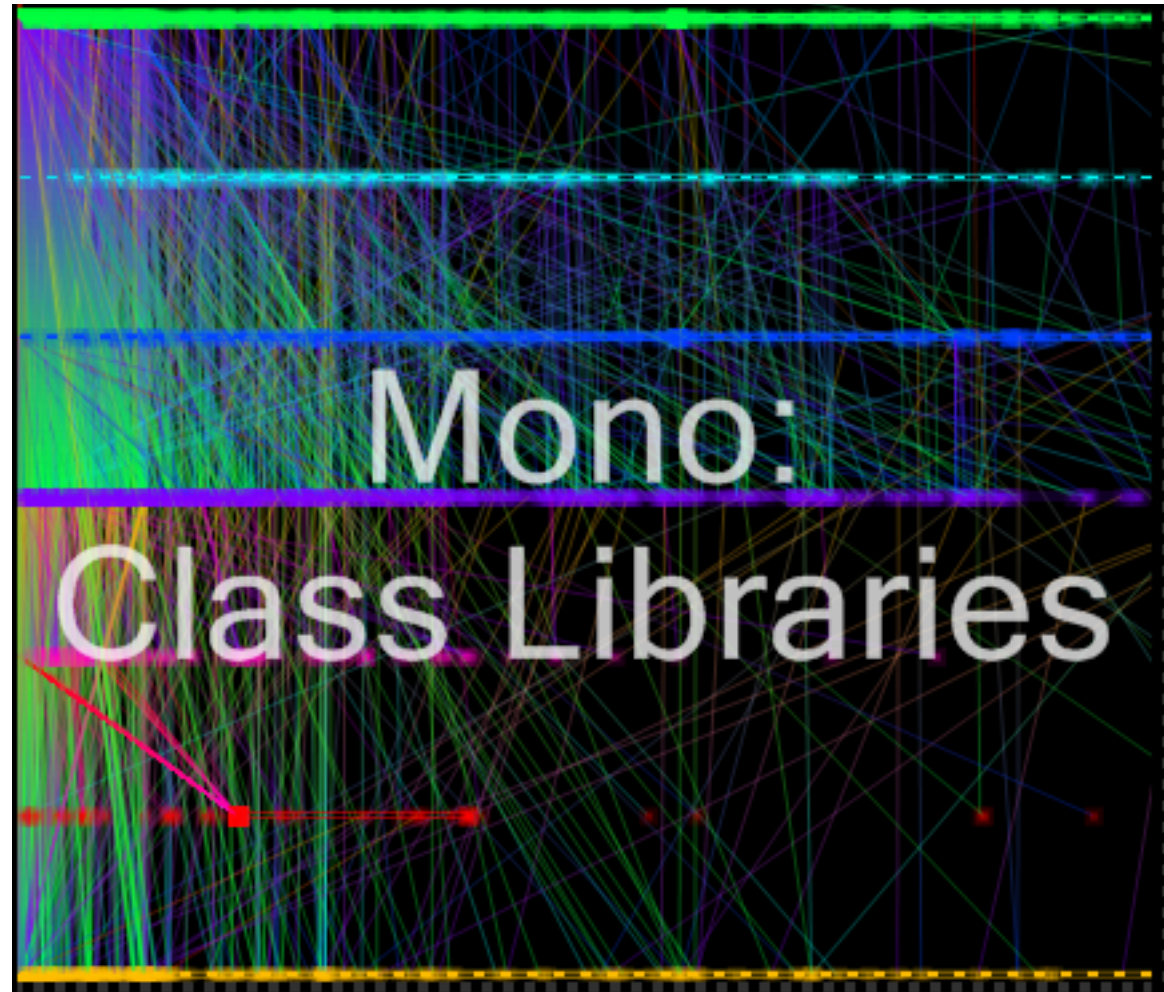
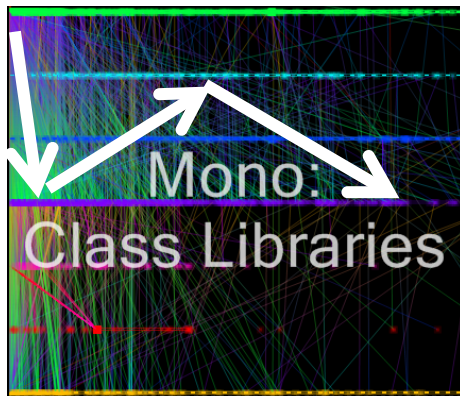


(例) 折れ線表現で見える特徴

□ 安定パターン



□ 手戻りパターン





大量の時刻パタンの可視化

ChronoView

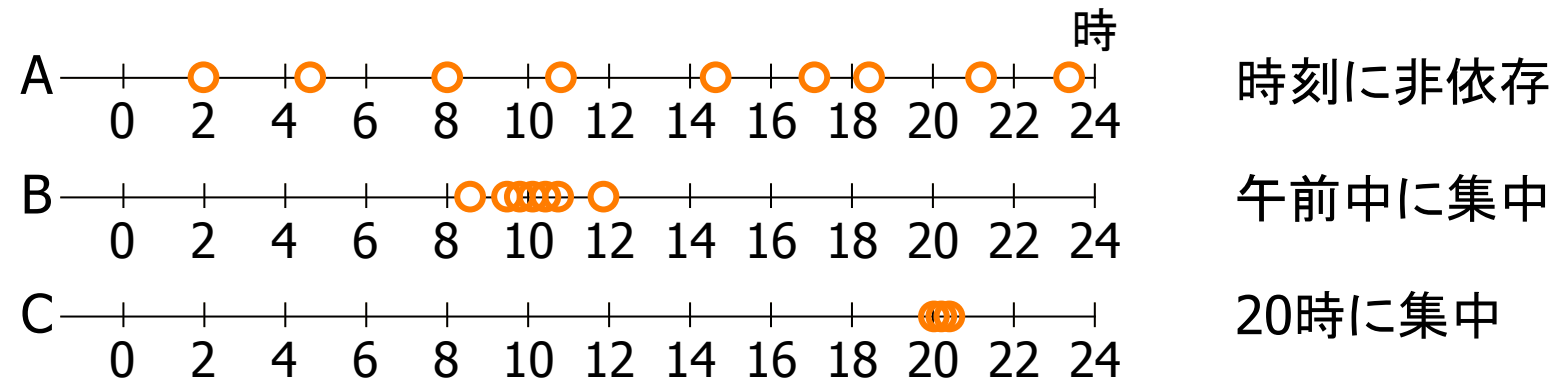
購買履歴データ(某研究室)

- 2006年6月～2007年8月 5308件の購買記録
- 購入者数は25人 販売商品377種類

購入日	購入時刻	購入者	商品名	カテゴリ	価格
2006/6/2	22 32 24	Szyu	ホームラン軒鶏ガラ醤油	インスタント食品	100
2006/6/5	20 42 47	M asa	午後の紅茶 ミルクティー	清涼飲料水	100
2006/6/5	20 37 35	N aga	サッポロー 番塩らーめん	インスタント食品	100
2006/6/5	20 31 20	K S th	スーパーカップ 1.5倍濃 コクとんこつ	インスタント食品	100
2006/6/5	19 39 50	Yuki	昔ながらのソース焼きそば	インスタント食品	100
2006/6/5	17 59 05	Shu j	カップヌードルカレー	インスタント食品	100
2006/6/5	16 50 14	M hro	爽健美茶	茶	100
2006/6/5	16 49 36	M hro	コアラのマーチチョコレート	菓子類	100
2006/6/5	15 09 38	Shu j	B l endy	コーヒー	100
2006/6/5	14 42 52	Jacq	深い味わい若武者	茶	100
2006/6/5	14 35 13	S n j i	コカコーラ	清涼飲料水	100
2006/6/5	13 39 50	Kazu	午後の紅茶ストレートティー	清涼飲料水	100
2006/6/5	12 52 02	N aga	爽健美茶	茶	100
	12 22 17	S n j i	爽健美茶	茶	100

時刻パターン

□ タイムスタンプの集合の特徴

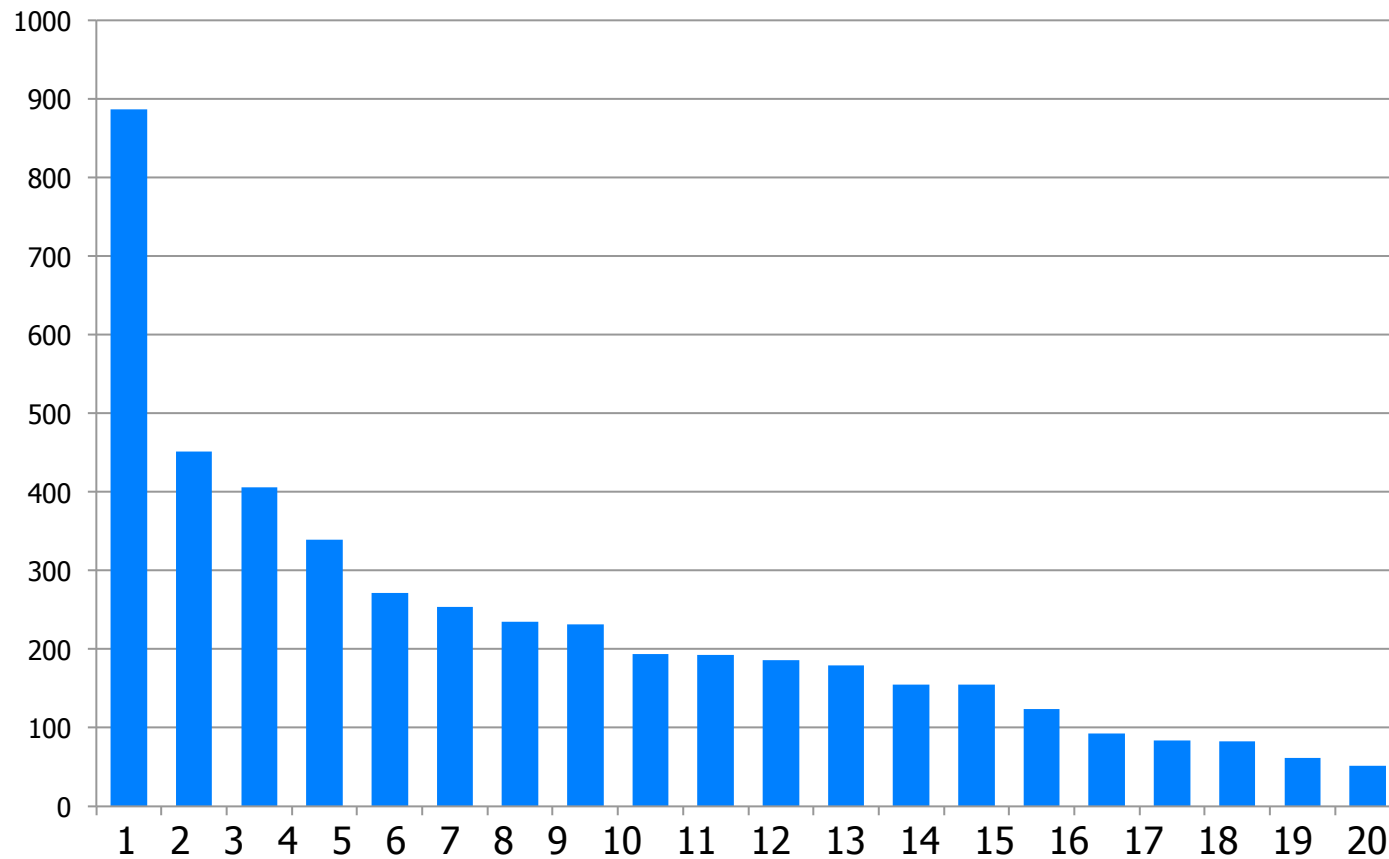


□ イベントデータの時刻パターンを捉えることは重要

- 例) 購買履歴: ミルクティーはいつ売れたか?
- 例) インシデント・ログ: A社サイトはいつアタックされたか?

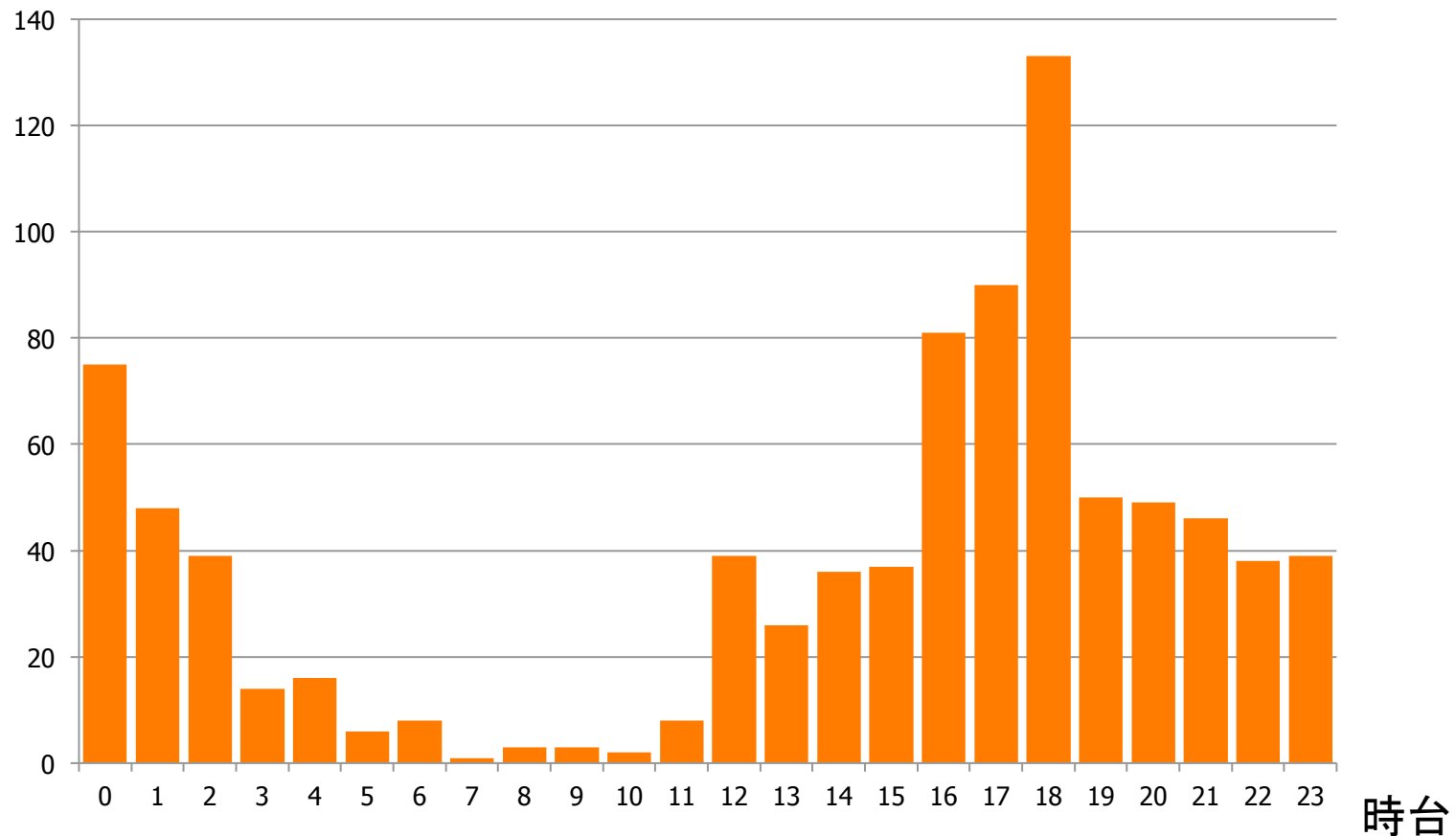
□ 関心は、個々の時刻にではなく、時刻パターンにある。

商品別売上数(上位20)



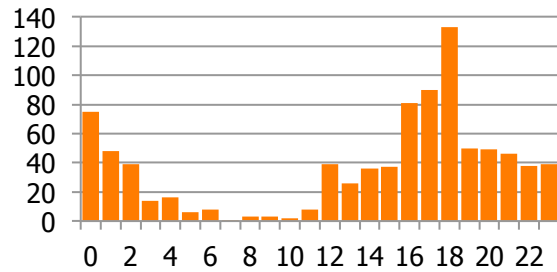
売上時刻パターン

「うまい棒」の時間帯別売上頻度

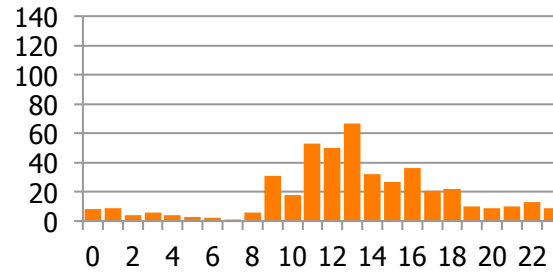


商品別売上時刻パターン

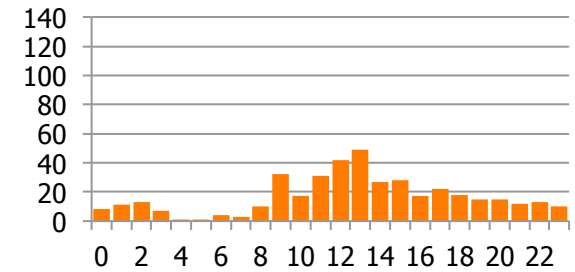
うまい棒



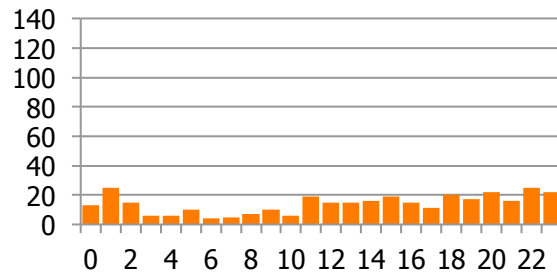
ミルクティー



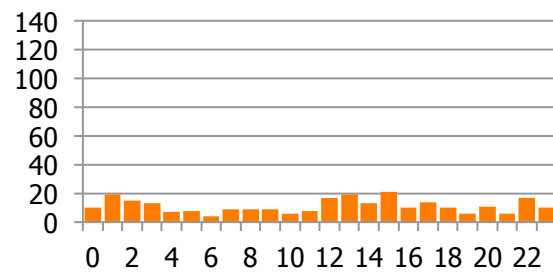
お茶



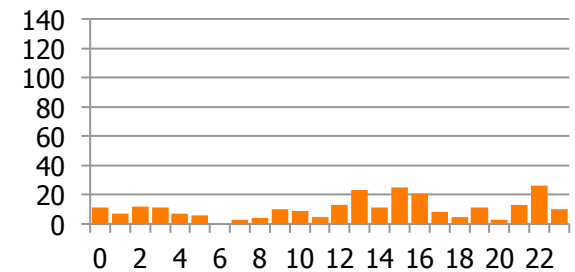
野菜ジュース



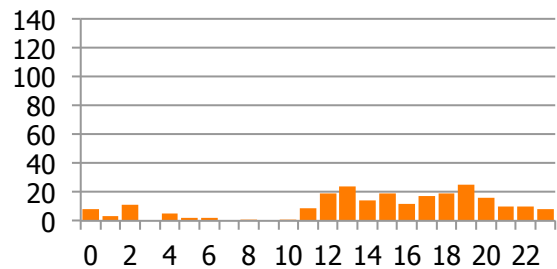
ビタミンドリンク



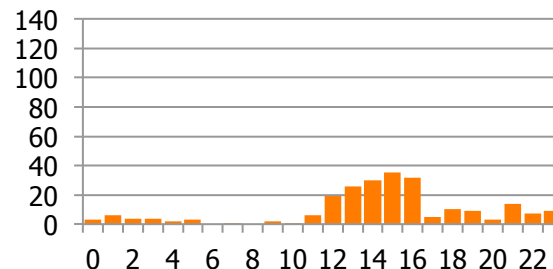
コーヒー



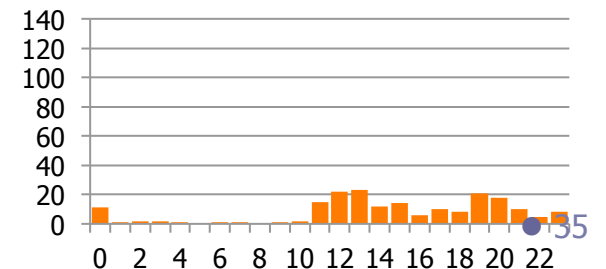
ラーメン



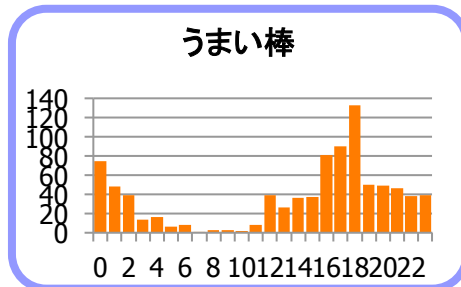
コココーラ



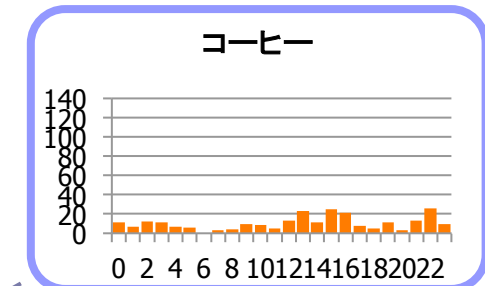
焼きそば



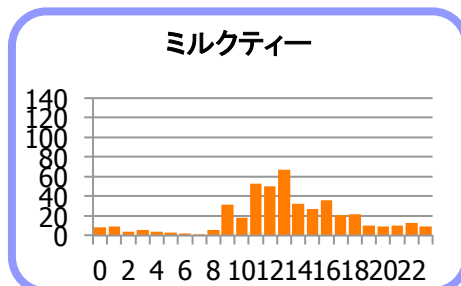
時刻パタンを位置で表現



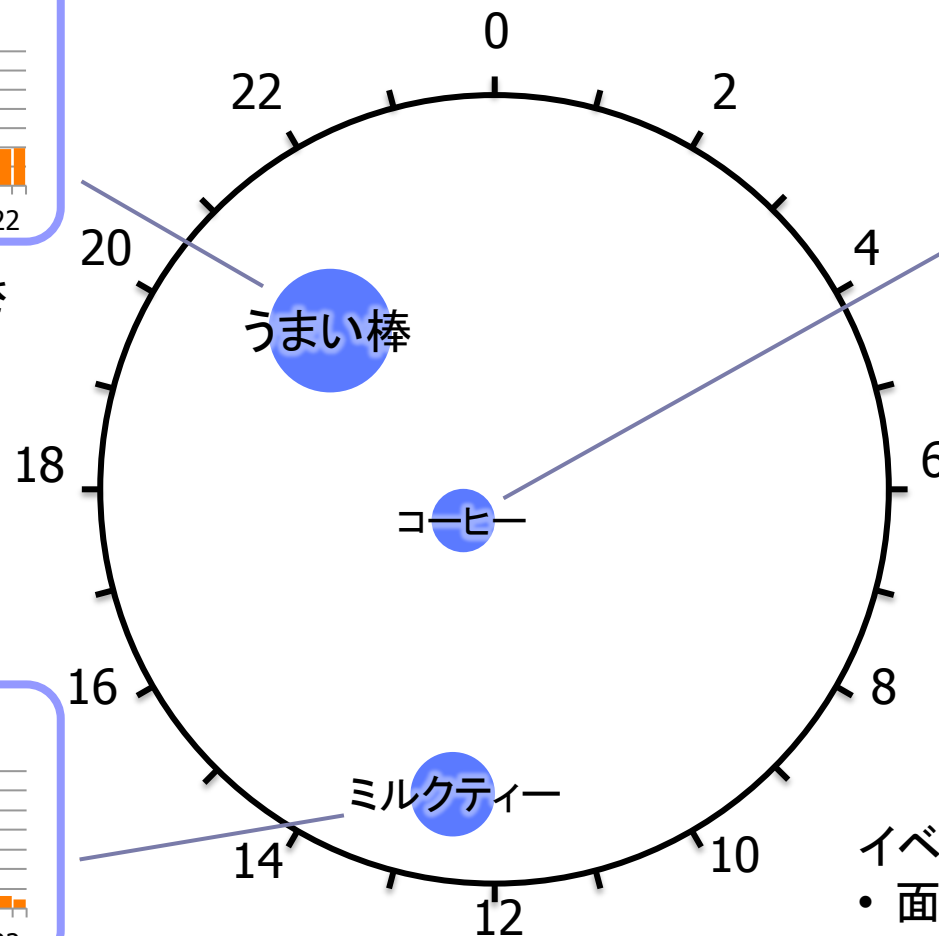
夕方から深夜にたくさん売れる



時間にあまり依存せずに売れる



昼食時を中心に売れる

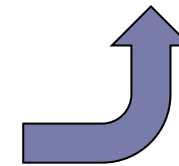
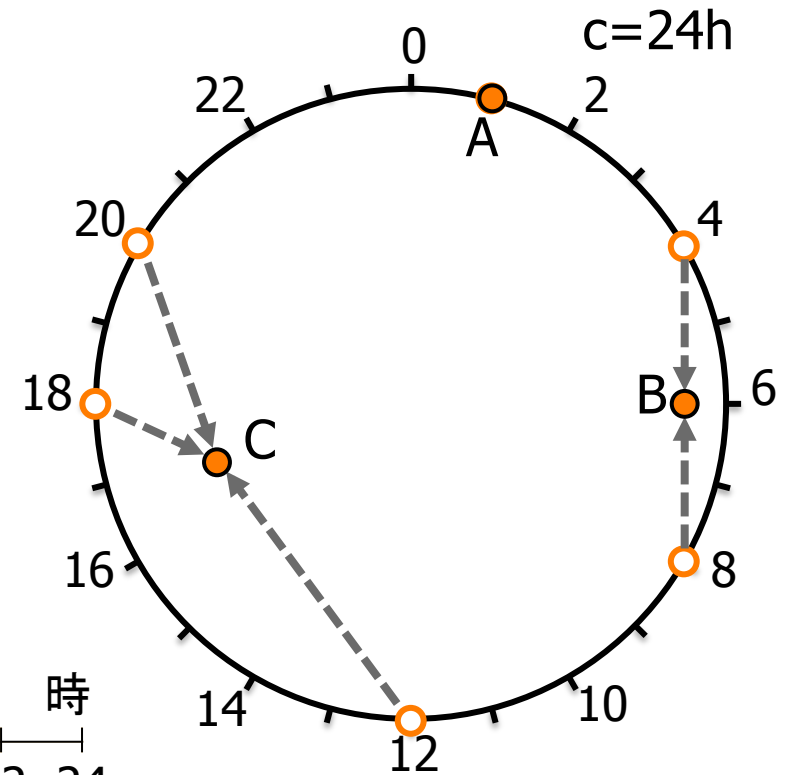
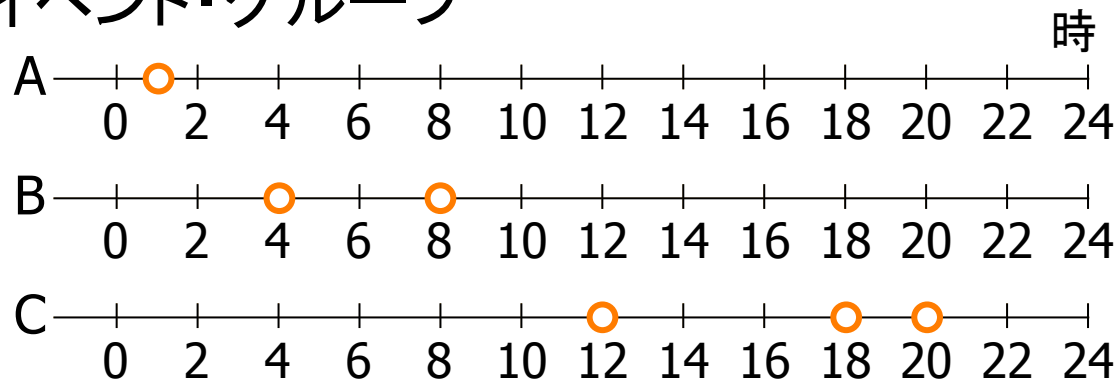


- イベントを円に集約。
- 面積が頻度を表す。
 - 位置が時刻パタンを表す。

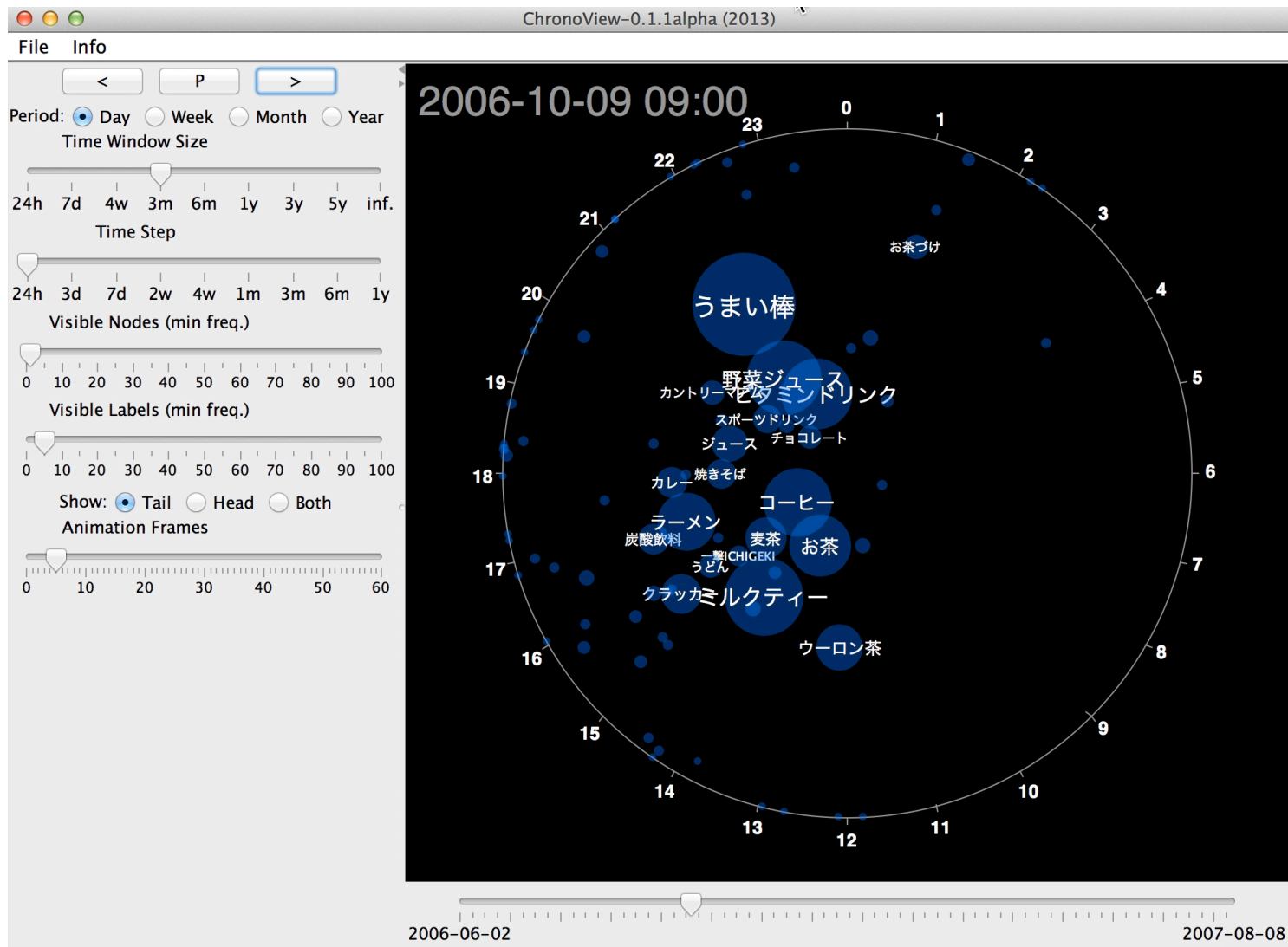
ChronoViewにおける時刻の表現

- アナログ時計の文字盤のように、円周上の位置で時刻を表す。
- 各イベント・グループの発生時刻を集計し、時刻の重心に配置。

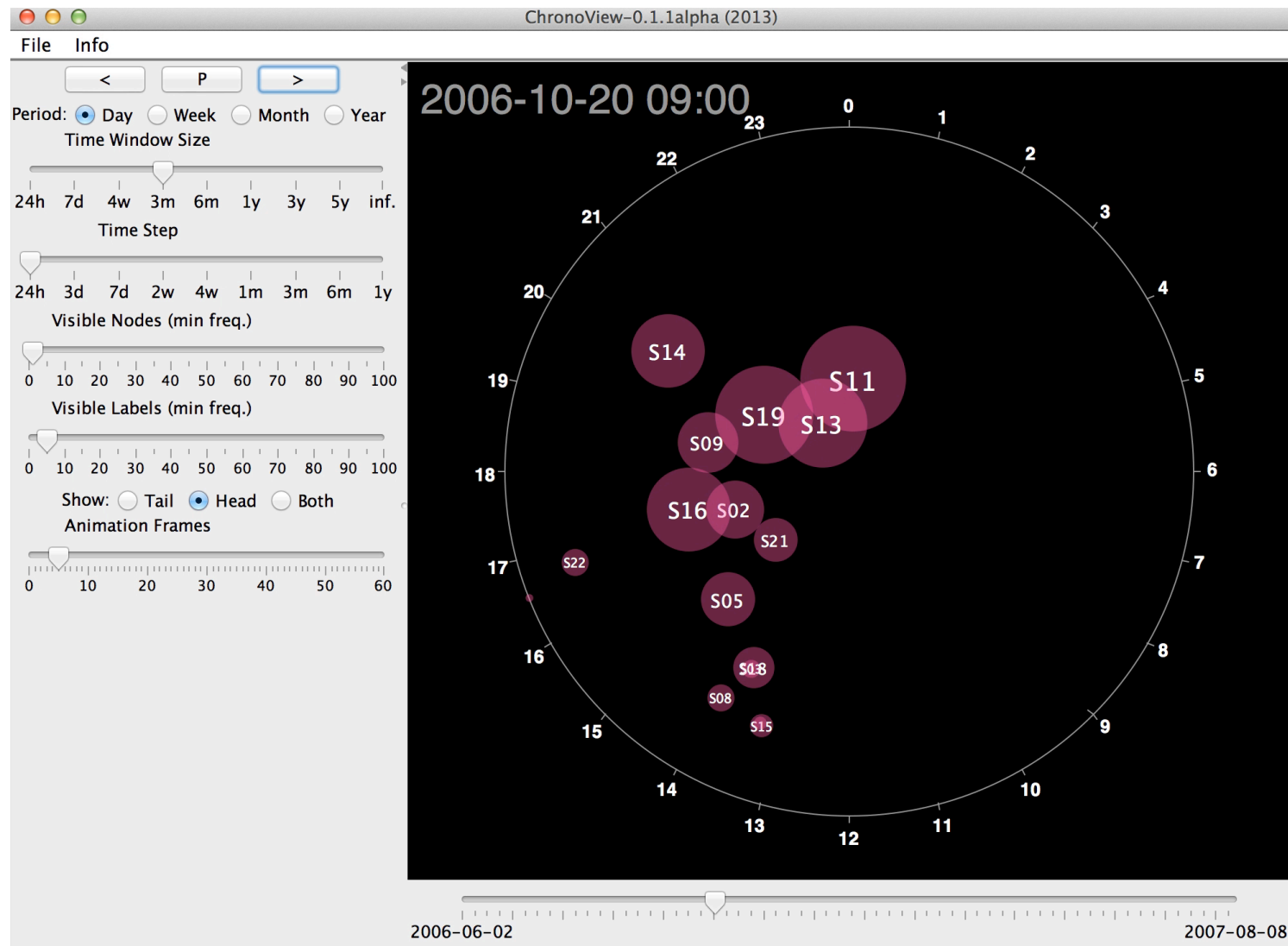
イベント・グループ



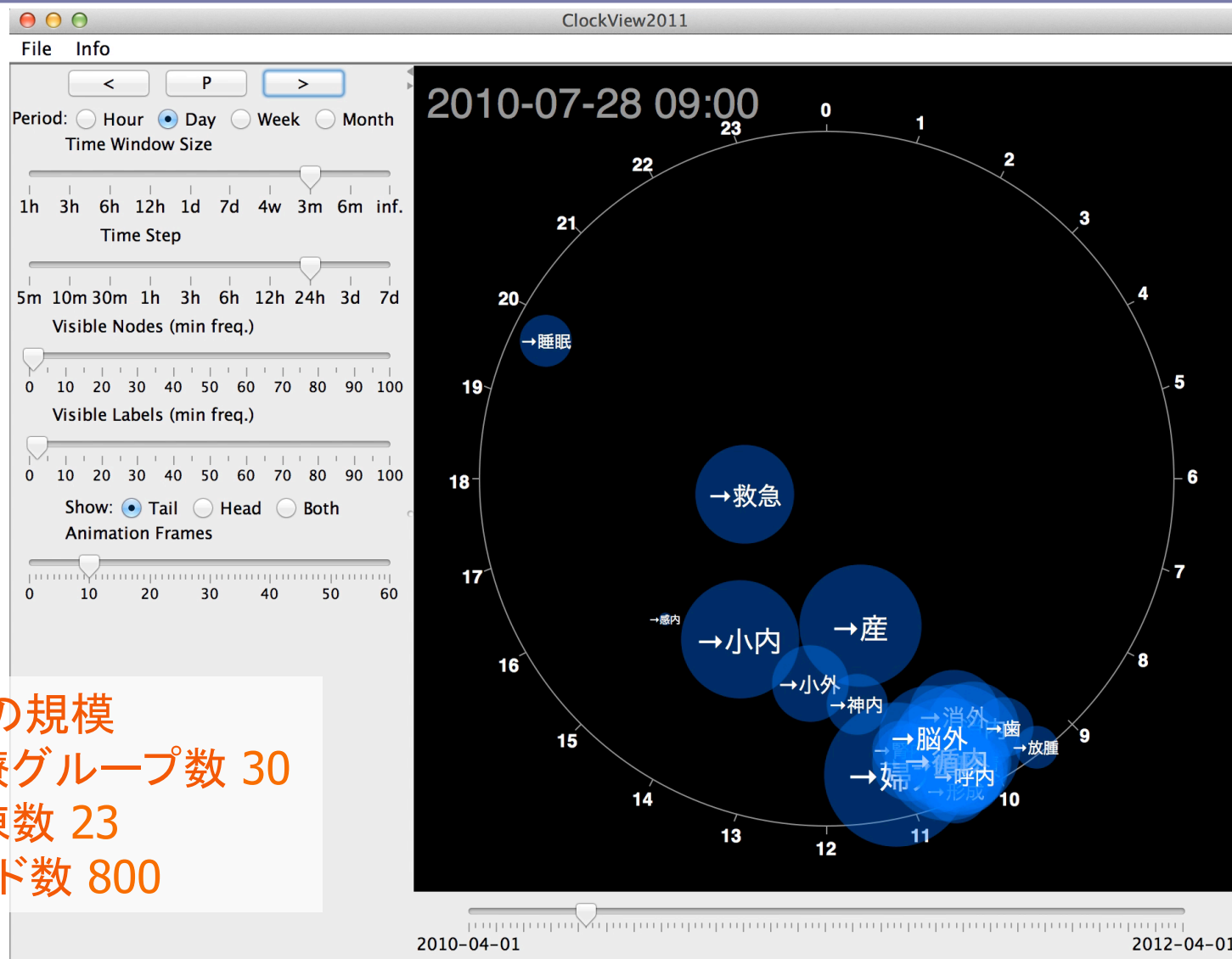
商品毎の販売時刻



顧客毎の購買時刻



患者の行動(筑波大学付属病院) 診療グループ別入院時刻パターン



病院の規模

- 診療グループ数 30
- 病棟数 23
- ベッド数 800



大規模な人の移動の集約型表現

Amoeba

人の流れデータとは

□ 人々の位置を時系列的に記録したデータ

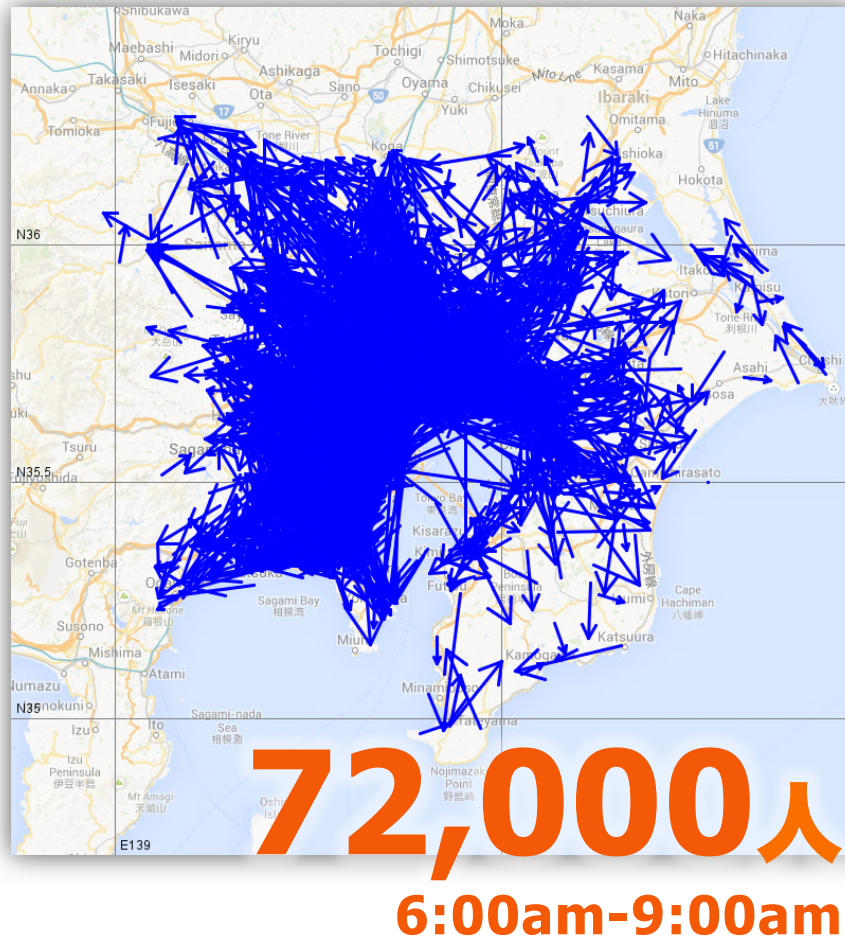
ID	時刻	緯度・経度
50	2014/04/05 13:00	36.110915, 140.099898
50	2014/04/05 17:00	36.094774, 140.098367
51	2014/04/05 13:00	36.098233, 140.105161
:	:	:

□ 位置の情報は様々な場面で記録されている

- 位置データをつけたSNSへの投稿
- 交通系ICカードの利用ログ
- ...

(東京大学空間情報科学研究センターの人の流れプロジェクトデータを利用)

項目型表現による可視化



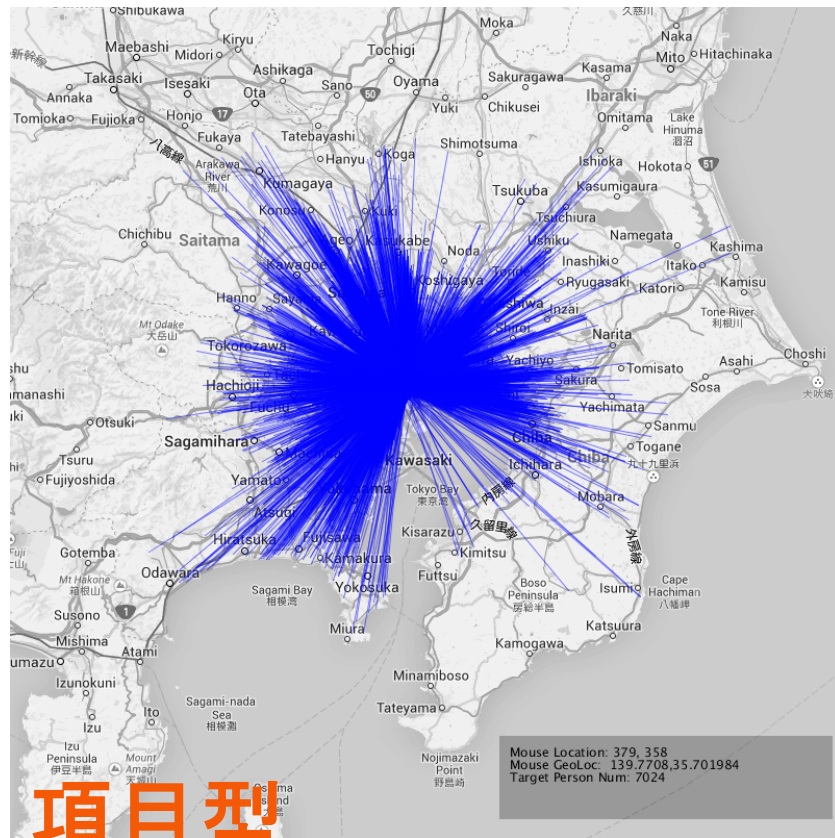
□ 始点と終点を線で接続

- 視覚的混雑により様々な情報が隠れている
- 中心部の向きは？
- 短い移動はある？

□ 項目型表現で大規模データを表すのは難しい

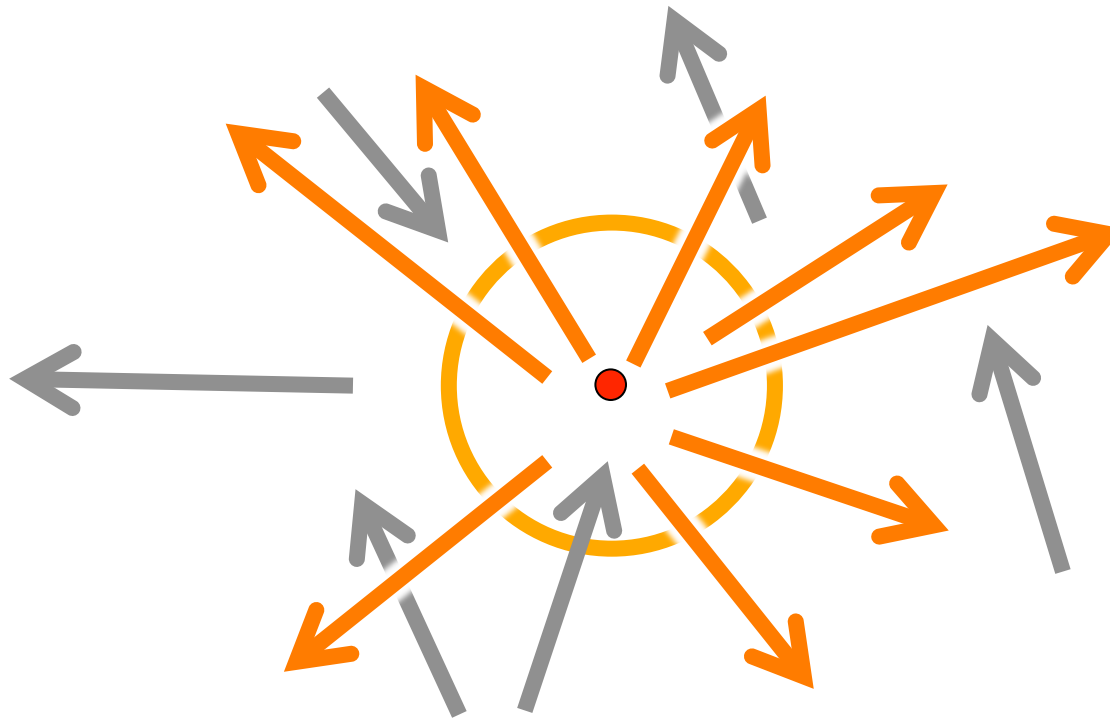
一地点近傍を起点と移動の可視化

秋葉原駅近傍(半径3km)を起点とする、約7000名、3時間の移動



Amoebaの描き方 始点による集約

- 始点がある近傍に含まれる移動を抽出する



Amoebaの描き方

移動方向による集約

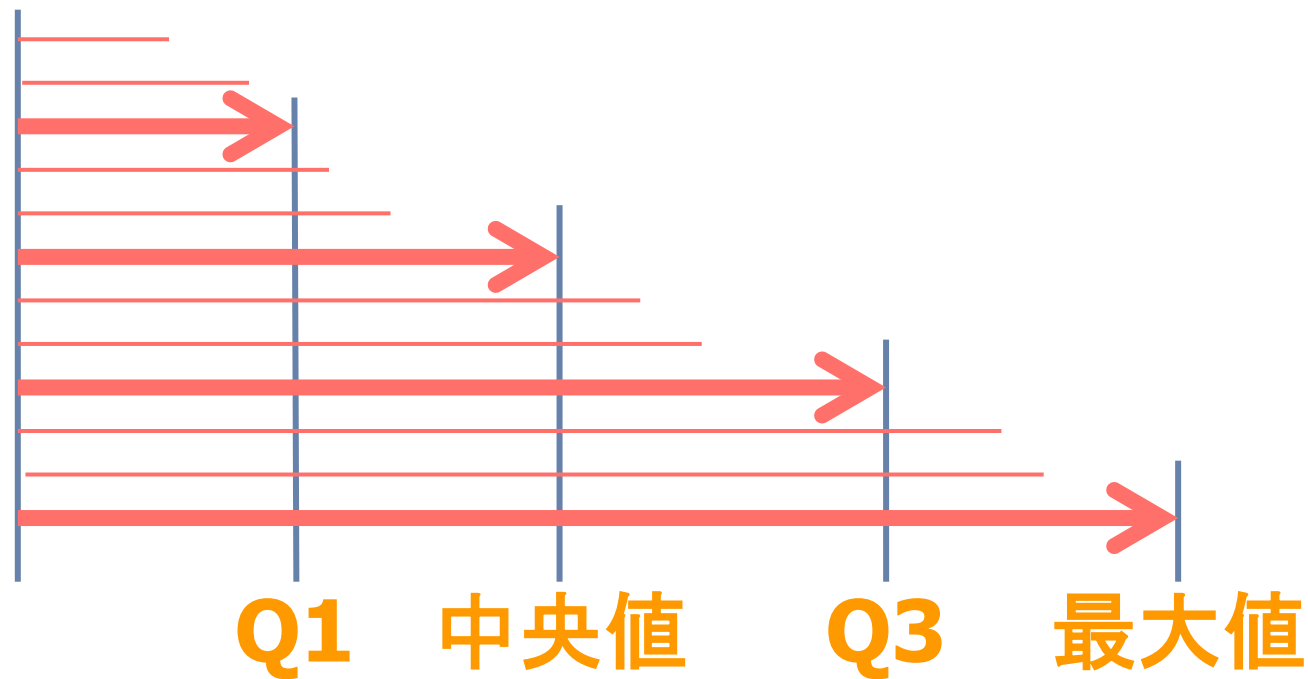
- 抽出した移動を、移動方向により k グループに分割



Amoebaの描き方

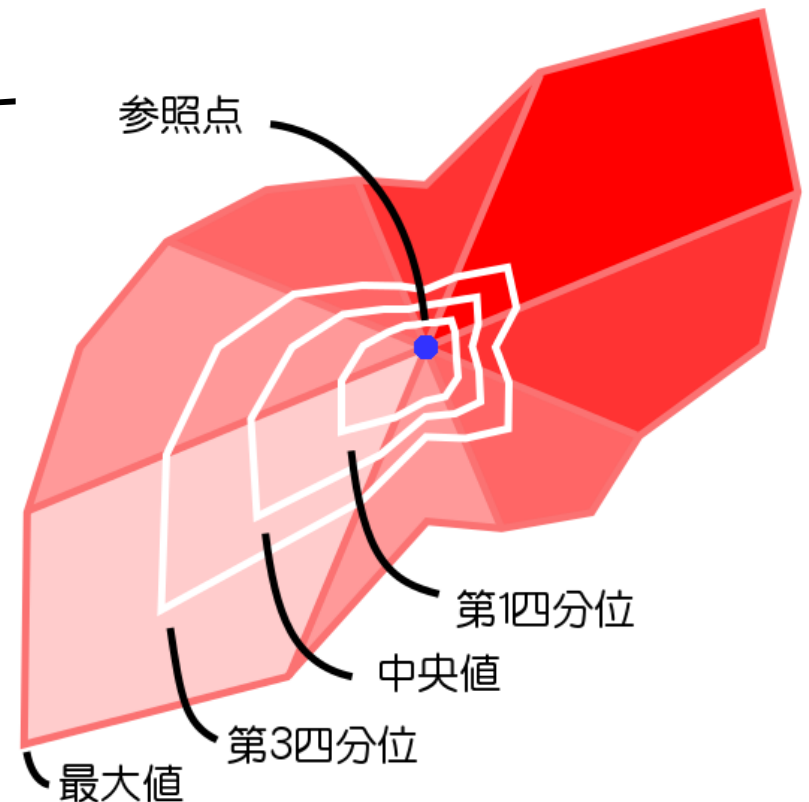
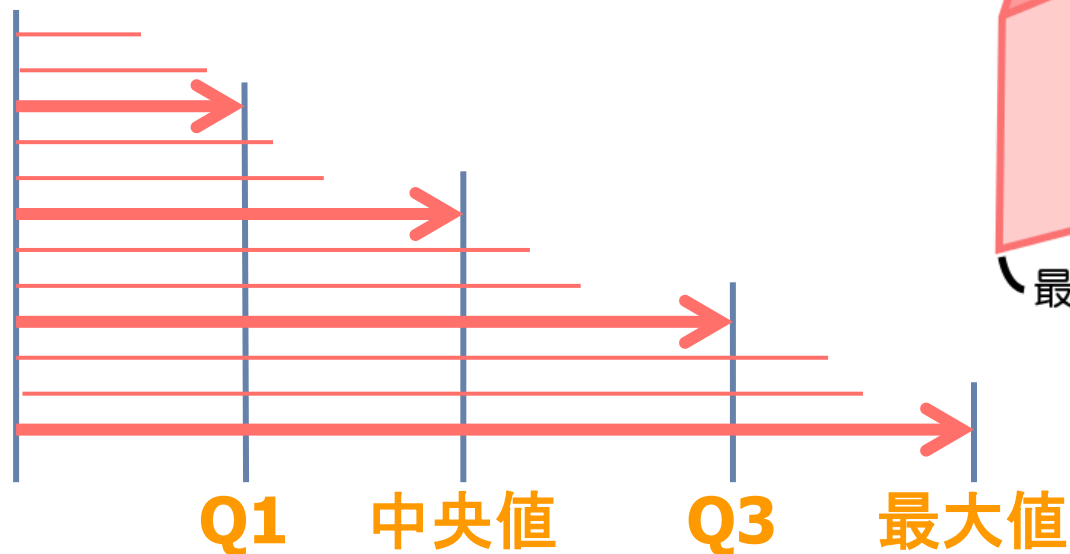
四分位値を求める

□ 移動距離の四分位値を求める



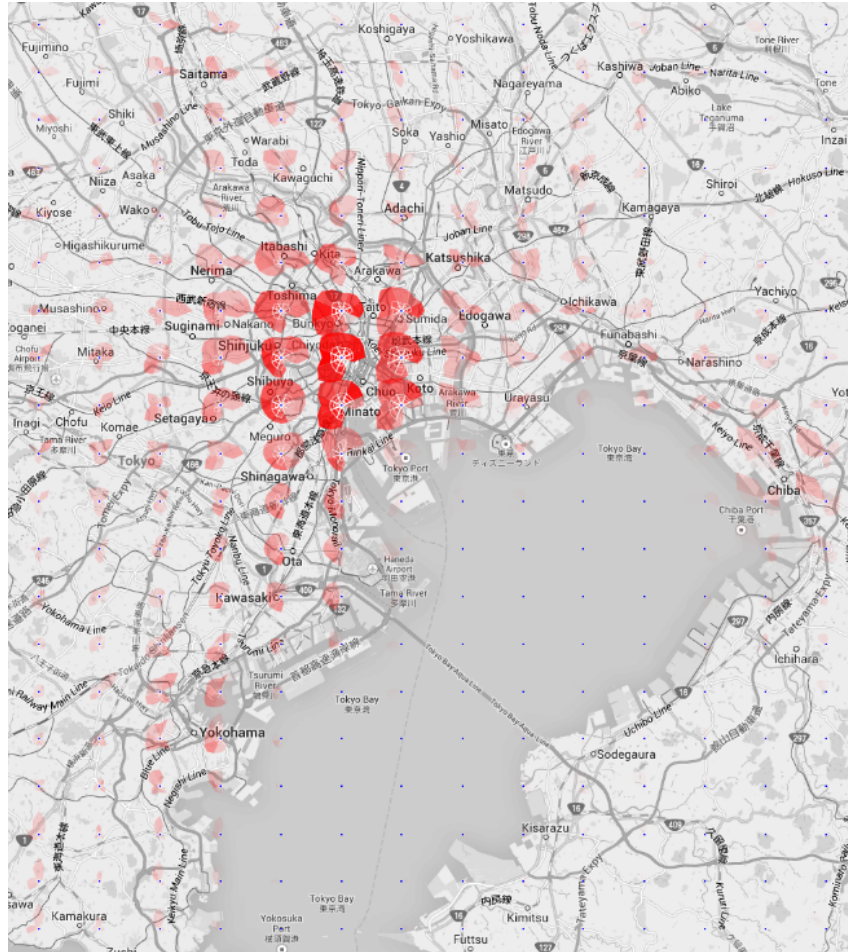
Amoebaの描き方 領域線を描く

- 各方向の四分位値を表す領域線を描き、移動人数を表す色で面を塗る



面積の影響を軽減するために面積比の移動人数を表現

Amoeba Colony 表現



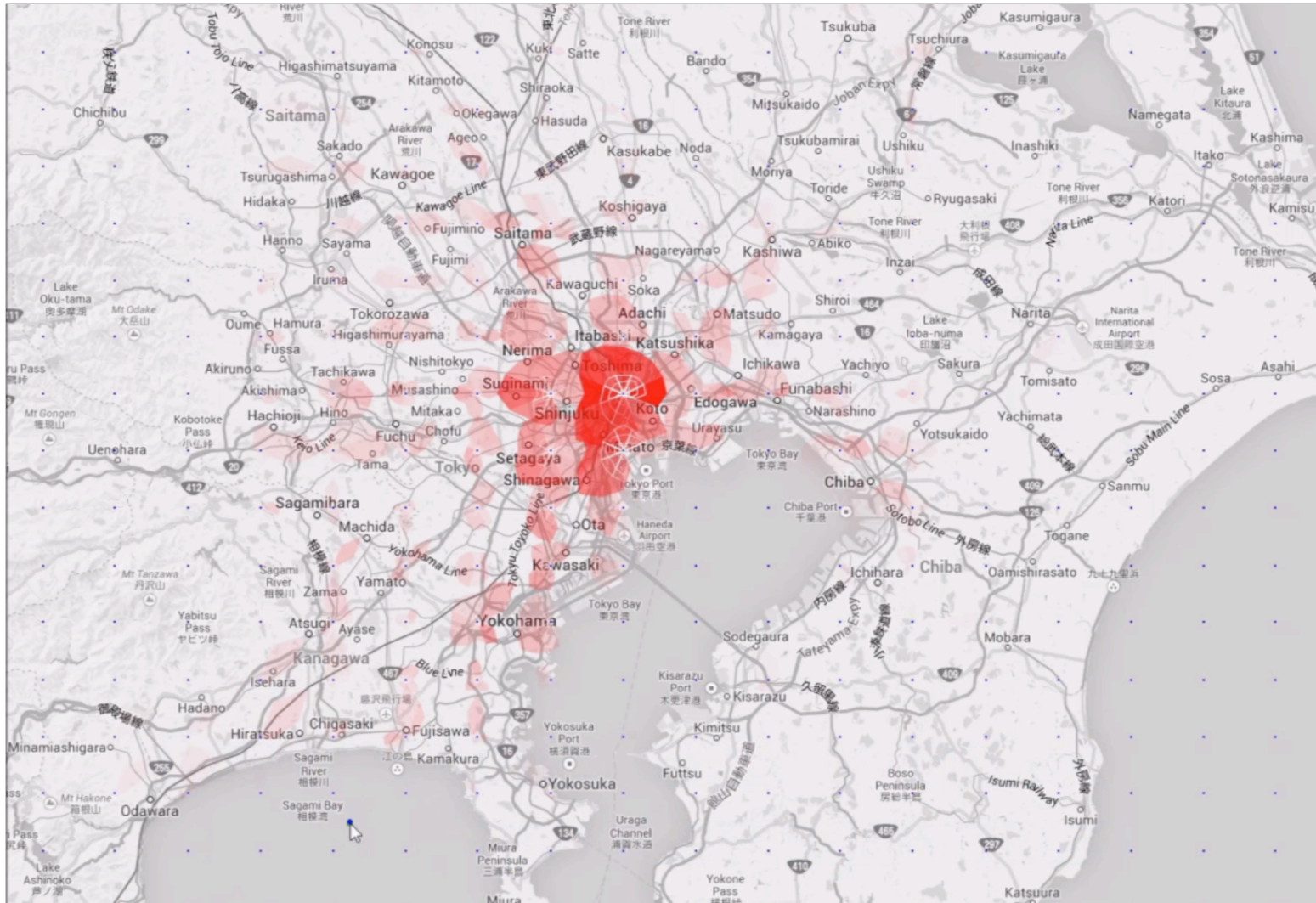
□ 多地点を始点とする移動を表現

- Amoeba はある一始点からの移動を表現。そのままでは多地点からの移動を表現できない。

□ 縮小したAmoeba をグリッド状に配置

- cf. Small multiples
- 各地からの移動の様子を一覧できる

デモ



まとめ

- 集約型表現とは
 - 複数のデータ項目を一つの図形で表現する
- 集約型表現の高度化への取り組み紹介
 - 高次元データ: Colored Mosaic Matrix
 - 多次元データをカテゴライズして着色を工夫
 - 複合データ: SaaG: Series at a Glance
 - Treemapに時系列データの視覚的表現を埋め込み
 - 時刻パターン: ChronoView
 - 時刻集合を平面上の位置で表現
 - 移動データ: Amoeba
 - 移動数を方向と距離(四分位)で集約し色で表現

参考文献

□ Colored Mosaic Matrix

- Hiroaki Kobayashi, Kazuo Misue, Jiro Tanaka, Colored Mosaic Matrix: Visualization Technique for High-Dimensional Data, In Proceedings of 17th International Conference Information Visualisation (iV2013), pp. 378-383, London, UK, July 15-18, 2013.

□ SaaG

- Kazuo Misue and Seiya Yazaki, Panoramic View for Visual Analysis of Large-scale Activity Data, Business Process Management Workshops, BPM 2012 International Workshop Revised Papers, Lecture Notes in Business Information Processing, Vol. 132, pp.756-767, Springer-Verlag, Berlin Heidelberg, 2013.

□ ChronoView

- Satoko Shiroi, Kazuo Misue, Jiro Tanaka, ChronoView: Visualization Technique for Many Temporal Data, In Proceedings of 16th International Conference Information Visualisation (iV2012), pp.112-117, Montpellier, France, July 10-13, 2012.
- Kazuo Misue, Yuta Kanai, and Hideaki Takagi, Visualizing the Overview of Temporal Patterns of Patients' Activities, 2013 Workshop on Visual Analytics in Healthcare (VAHC2013), pp. 11-14, Washington, D.C., USA, Nov. 16, 2013.

□ Amoeba

- Yuuki Hyougo, Kazuo Misue, and Jiro Tanaka, Directional Aggregate Visualization of Large Scale Movement Data, Proceedings of 18th International Conference Information Visualisation (iV2014), pp. 196-201, Paris, France, July 15-18, 2014.

ChronoView プログラム公開中

□ 実行可能jarファイルを公開中

<http://www.cs.tsukuba.ac.jp/~misue/open/software/chronoview/>

ChronoView: Visualization of Many Temporal Patterns

[\[Examples\]](#) [\[Quick Start\]](#) [\[Download\]](#) [\[License\]](#)

Overview

ChronoView is a visualization technique for many event groups with time-stamps. Target data are events, incidents, activities, and so on. For example, incident log, sales history, and tweets on Twitter. ChronoView places each event group on a plane similar to the clock face of an analog clock. Feature efficiency for very

References

- [1] Satoko Shirai, Proceedings of 16th International Conference Information Visualization (IV2012, July 10-13, 2012, Montpellier, France), pp.112-117, 2012. [\[PDF\]](#)
- [2] Kazuo Misue, Yuta Kanai, and Hideaki Takagi, Visualizing the Overview of Temporal Patterns of Patients' Activities, Proceedings of the 2013 Workshop on Visual Analytics in Healthcare (VAHC2013, Nov. 16, 2013, Washington, D.C., USA), pp. 11-14, 2013. [\[PDF\]](#)

Examples

Visualization of Patients' Activities [2]

試用して感想・意見を下さい

Questions?



misue@cs.tsukuba.ac.jp
[@k3sue](#)