

Statistical Graphics & Visual Data Mining in the Medical Field

Jürgen Symanzik

Utah State University, Logan, UT, USA

*e-mail: symanzik@math.usu.edu

WWW: <http://www.math.usu.edu/~symanzik>



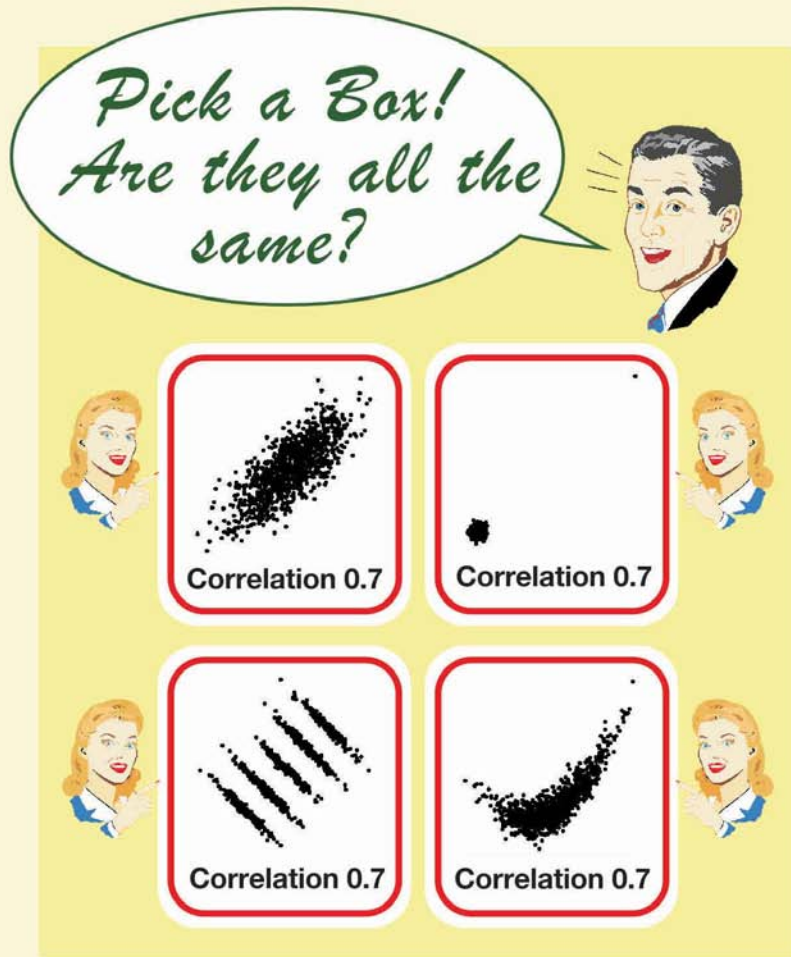
Contents

- Why Graphics for “Outsiders”?
- Exploratory Graphics for Actigraphy Data
- Carpal Tunnel Syndrome Data
- cgh Arrays and Gene Expression Data
- Conclusion

Interview with Andreas Buja*

- “ ... when I think back about what really may have had the most impact in what I did in the various labs that I worked, it’s graphics! You know whenever I made a striking picture, people actually went “aahh,” “wow,” “that’s great!”, “Why don’t we do more of this?” Pictures really, really speak. ...”

Inspiration for Statistical Graphics



American Statistical Association
Statistical Graphics Section
Poster Series (~2004)

<http://www.public.iastate.edu/~dicook/Sat.Stat.Graphics/posters.html>

SECTION ON
STATISTICAL
GRAPHICS



AMERICAN
STATISTICAL
ASSOCIATION

<http://www.amstat-online.org/sections/graphics/>

- **Example:**

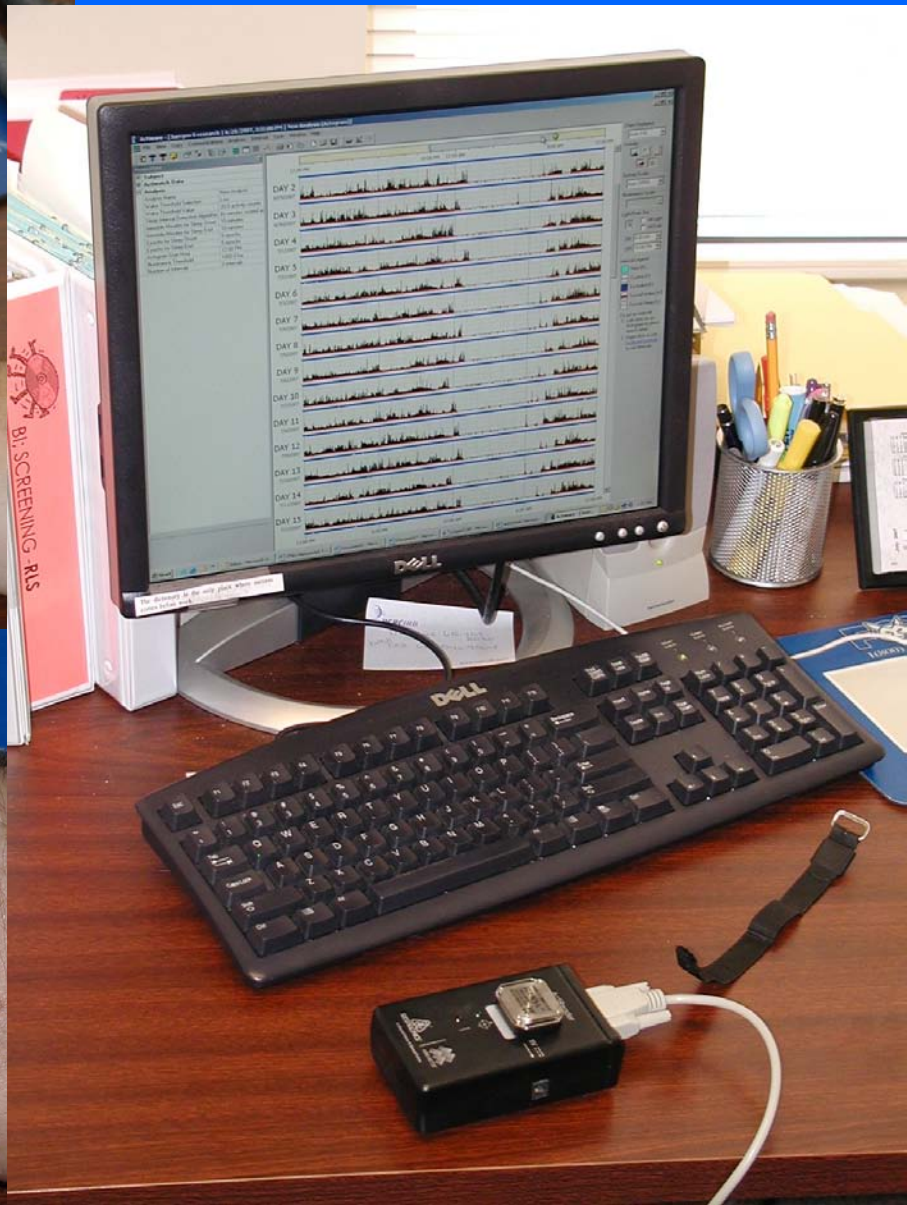
Exploratory Graphics for Functional Actigraphy Data in Sleep Medicine

- **References:**

**Symanzik, J., Shannon, W. (2008):
Exploratory Graphics for
Functional Actigraphy Data,
American Statistical Association,
CD, Forthcoming.**

Background (1)

- **Actigraphy**: emerging technology for measuring a patient's activity level continuously over time
- **Actigraph**: watch-like device (attached to the wrist or a leg) that uses an accelerometer to measure (human) movements (every minute or more often)

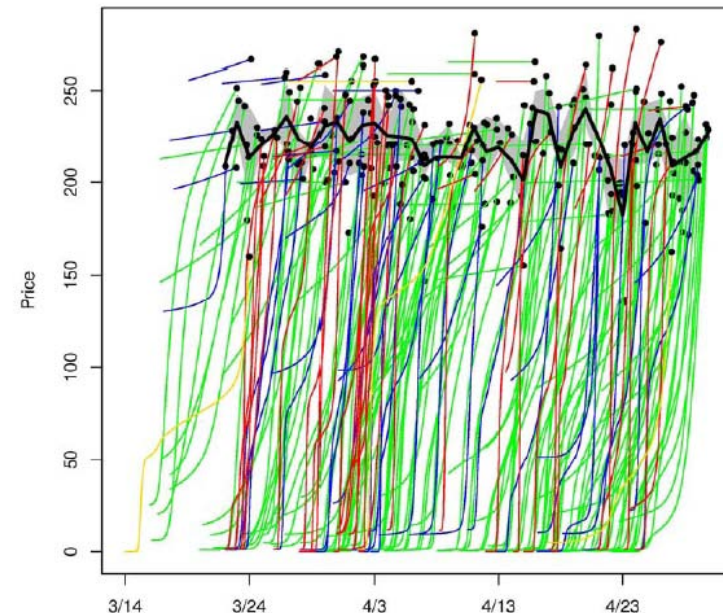


Background (2)

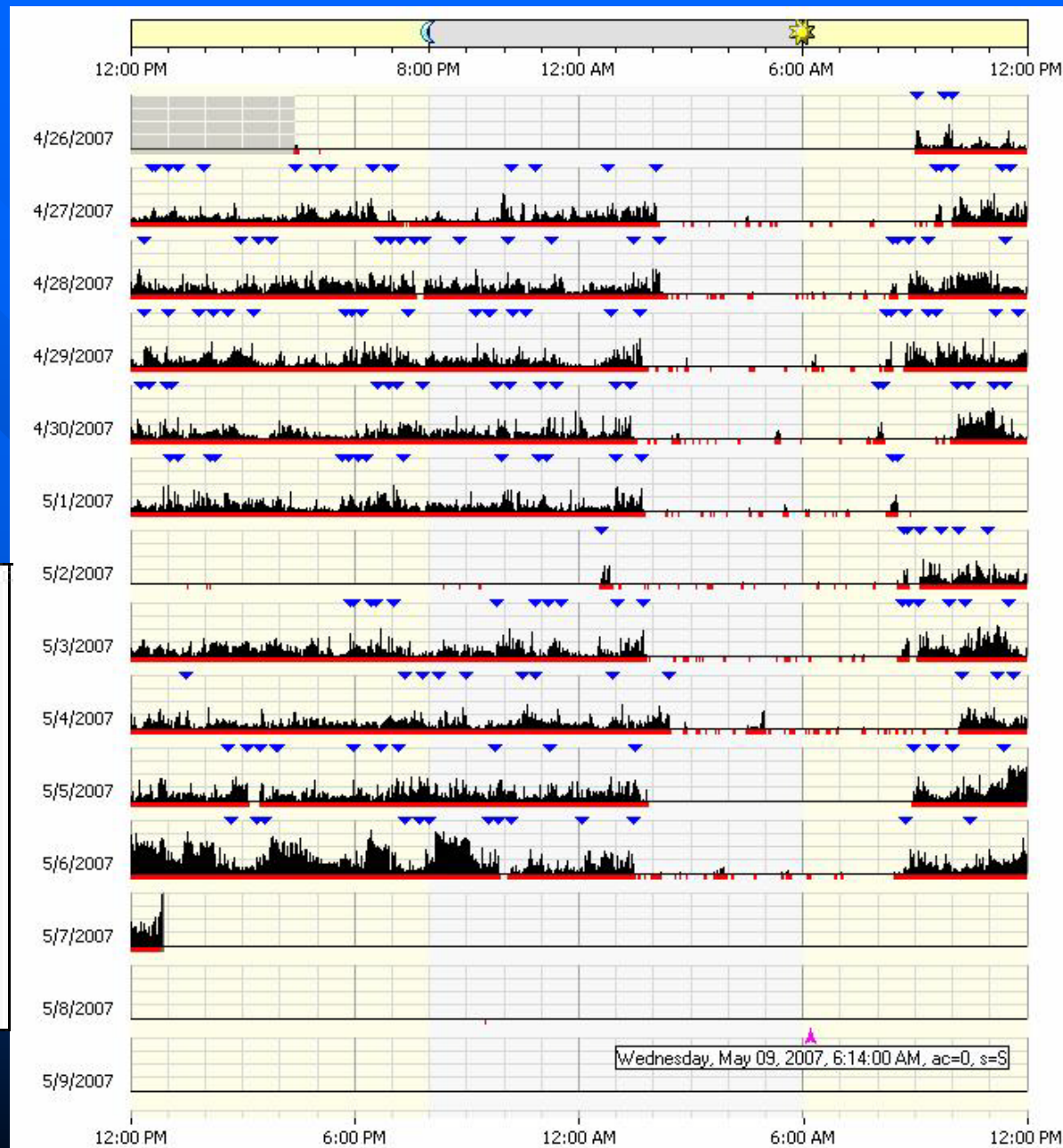
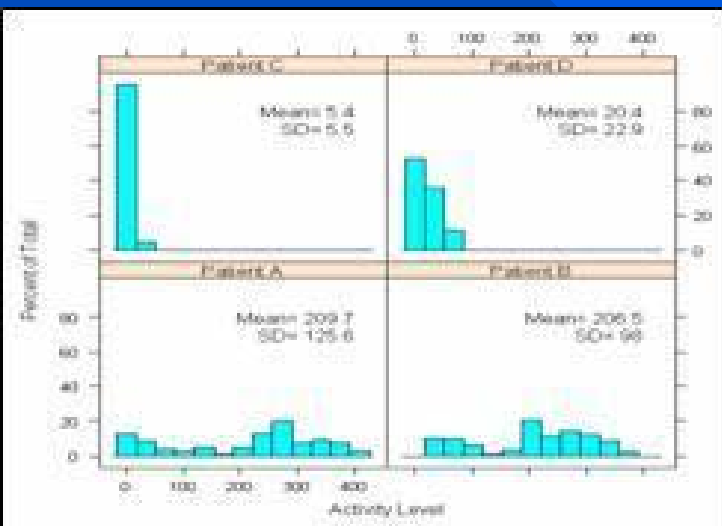
- **Analysis of Human Actigraphy Data:**
Useful for detecting sleep, for assessing insomnia and restless leg syndrome, for tracking recovery after heart attacks, and as an assessment tool for overall status of HIV patients
- Actigraphy Data can be best described as functional data

Visualization of Functional Data

- Very limited ! A rare example is:
Jank, W., Shmueli, G., Plaisant, C., Shneiderman, B.
(2008): Visualizing Functional Data with an Application to
eBay's Online Auctions, In: Chen, C., Härdle, W., Unwin,
A. (Eds.), Handbook of Data Visualization, Springer,
Berlin/Heidelberg, 873-898.
- Figure from
<http://www.smith.umd.edu/faculty/wjank/DIV-Berlin2006.pdf>
(page 30).

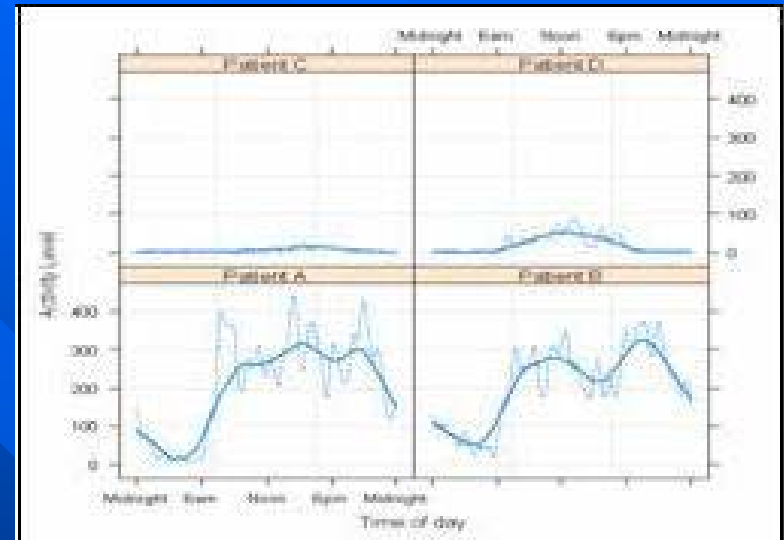


Current Visualization of Actigraphy Data

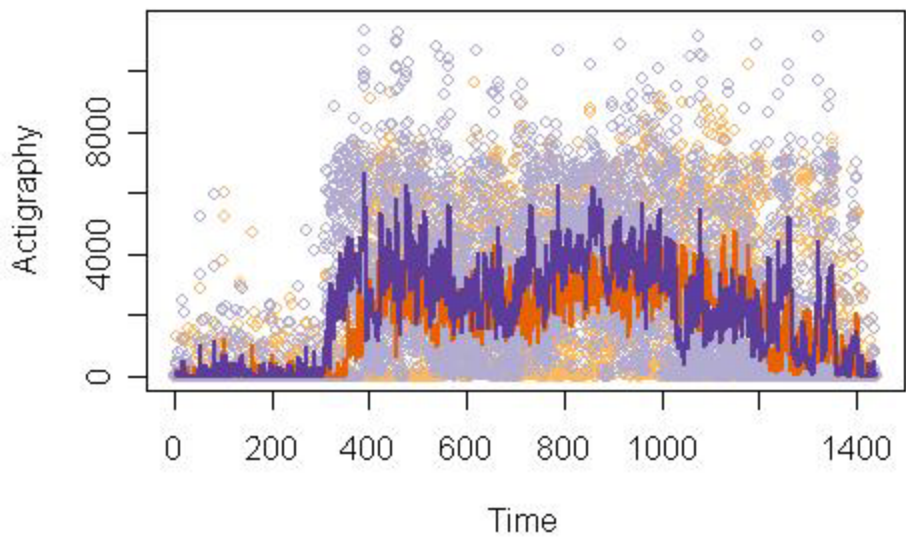


Suggested Future Visualization of Actigraphy Data (1)

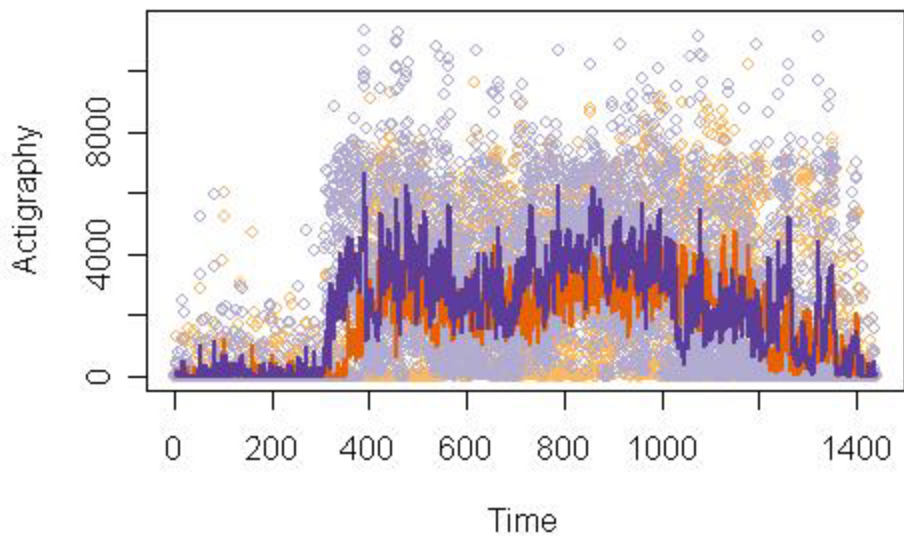
- Displays for
 - Raw data
 - Smoothed data
 - Averages etc.
 - Velocity (First Derivative)
 - Acceleration (Second Derivative)
 - Brushing & Linking
 - Cumulative Sums
- Example: 1 Subject
 - Orange: 5 Days at Baseline
 - Purple: 5 Days after 6 Months



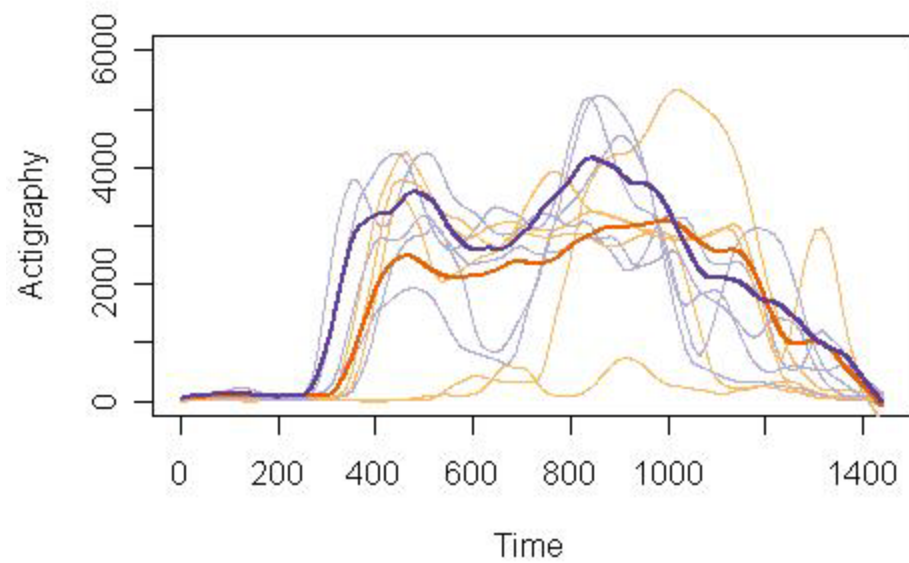
Raw Data



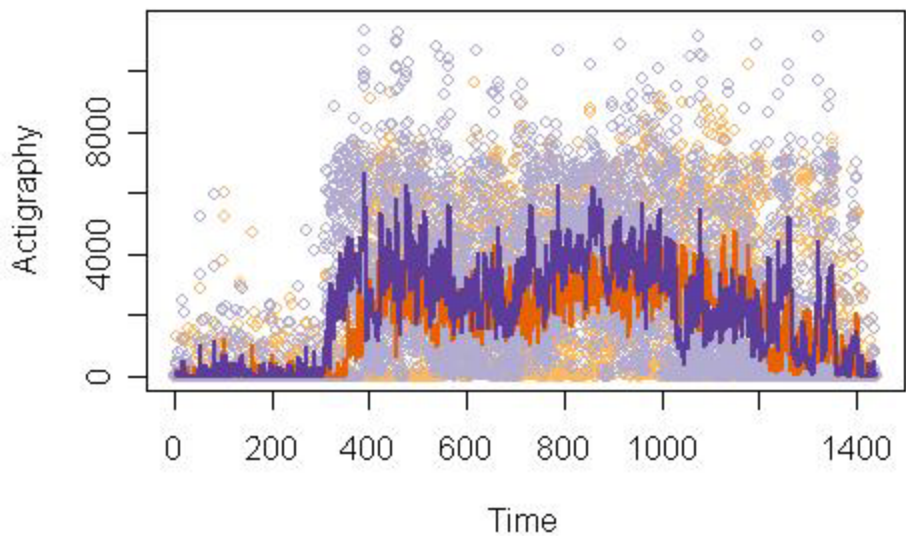
Raw Data



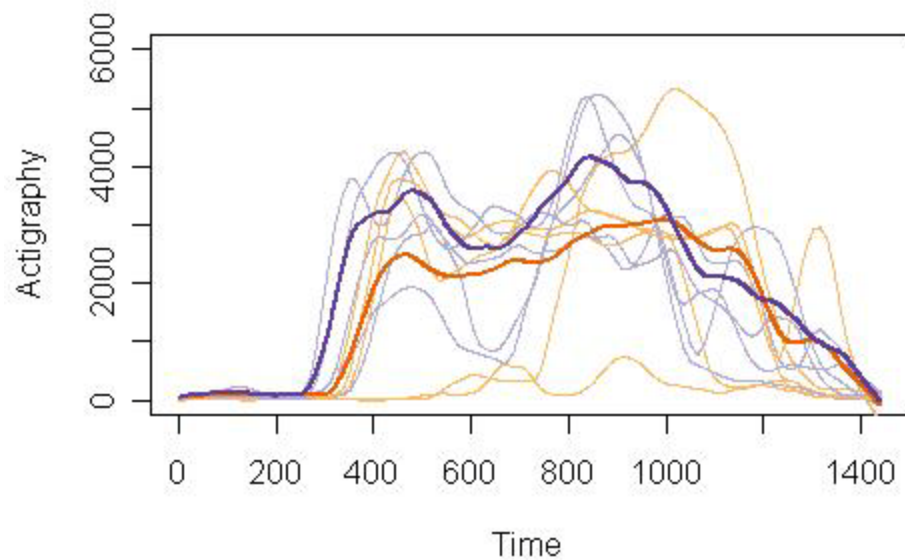
Smoothed Daily Data



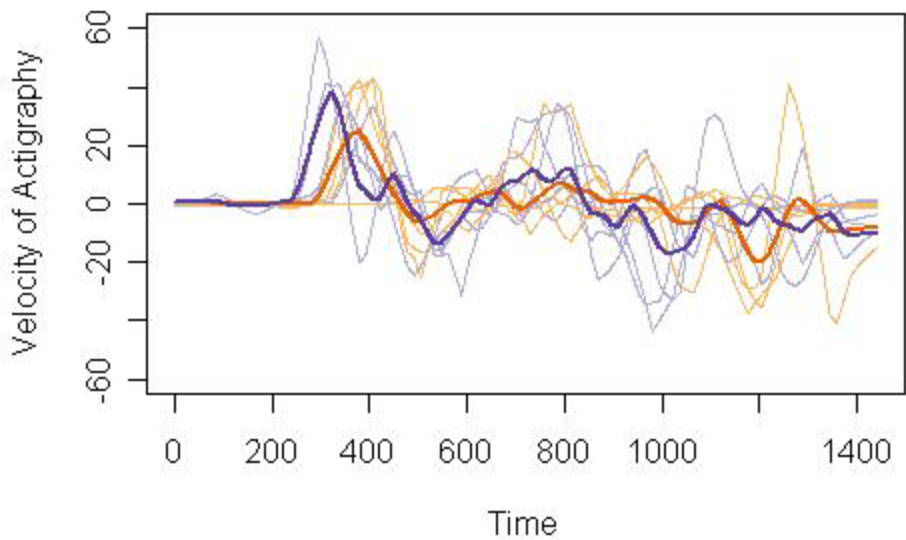
Raw Data

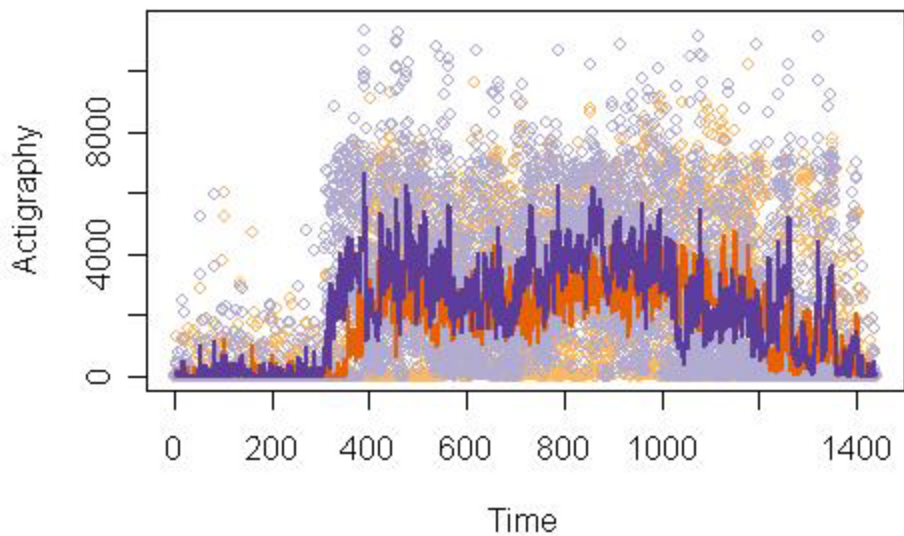
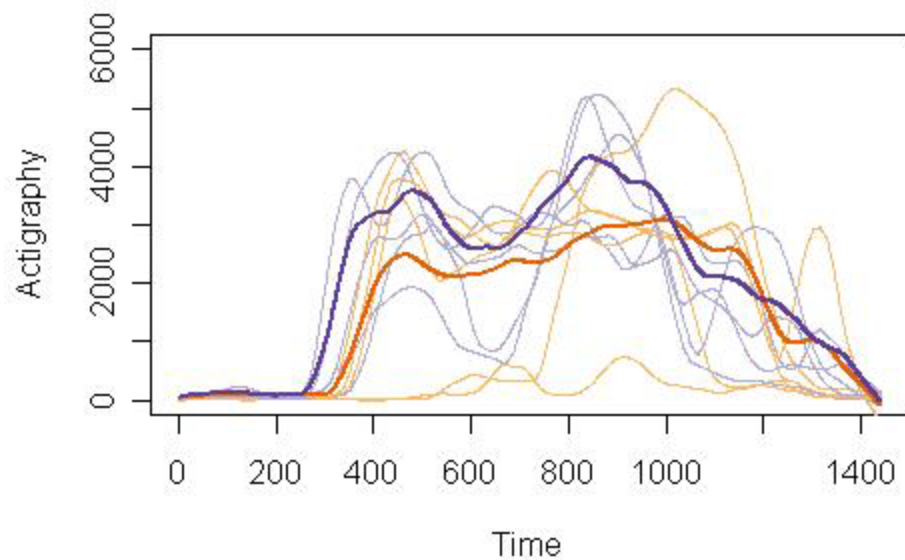
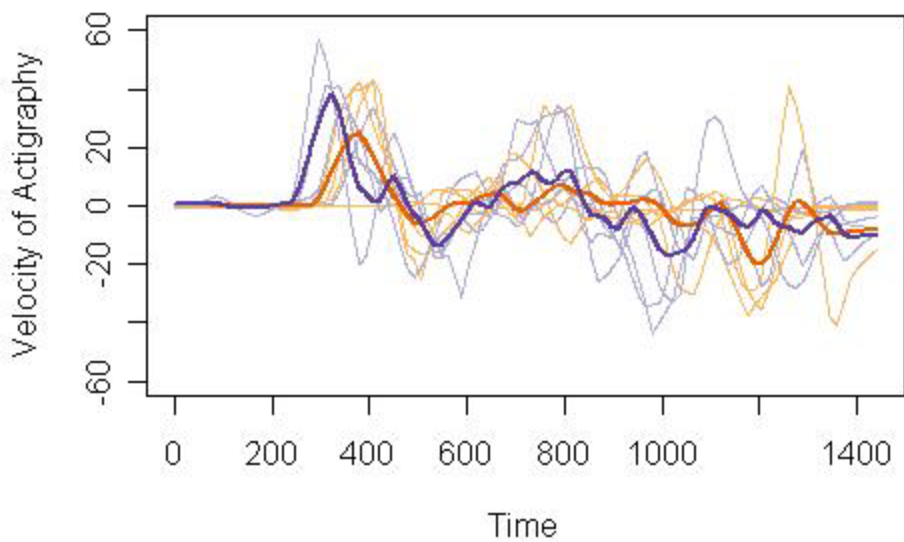
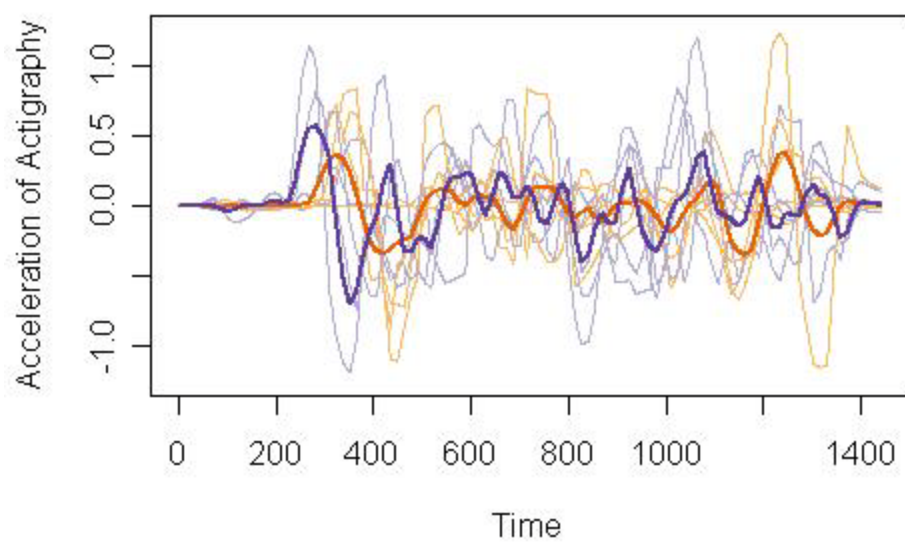


Smoothed Daily Data

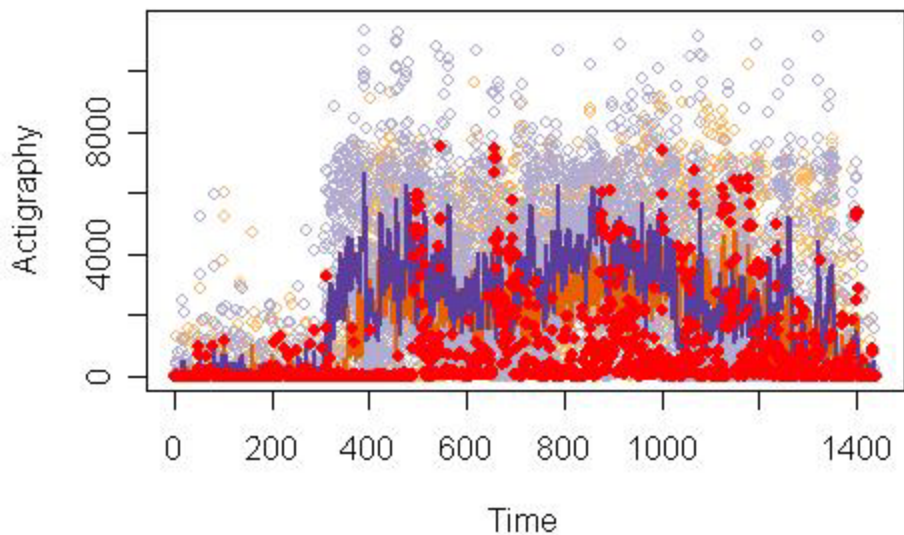


Velocity (First Derivative) of Smoothed Daily Data

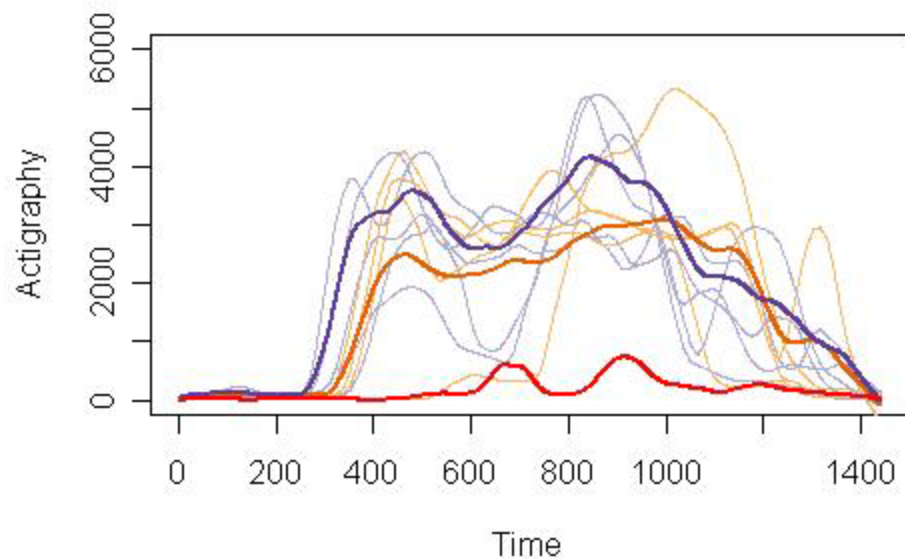


Raw Data**Smoothed Daily Data****Velocity (First Derivative) of Smoothed Daily Data****Acceleration (Second Derivative) of Smoothed Daily Data**

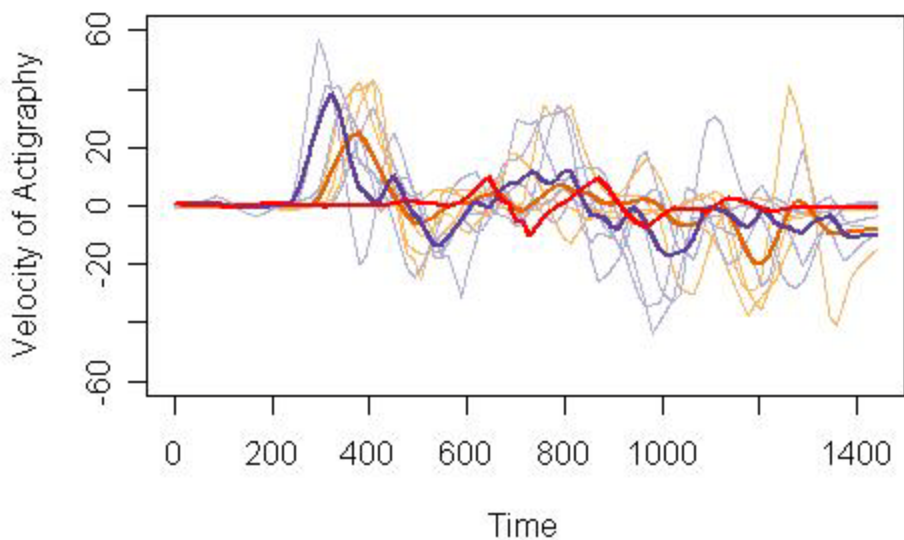
Raw Data (Base Day 3 Brushed)



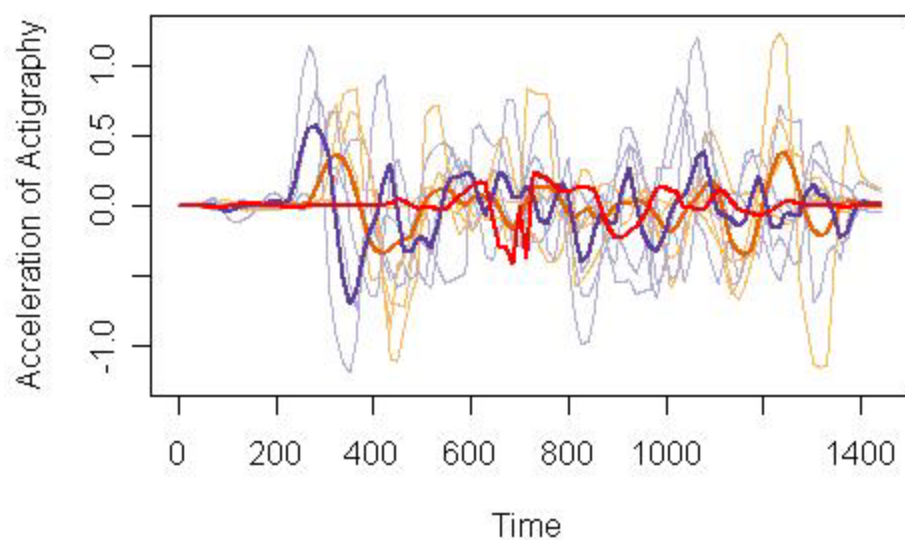
Smoothed Daily Data (Base Day 3 Brushed)



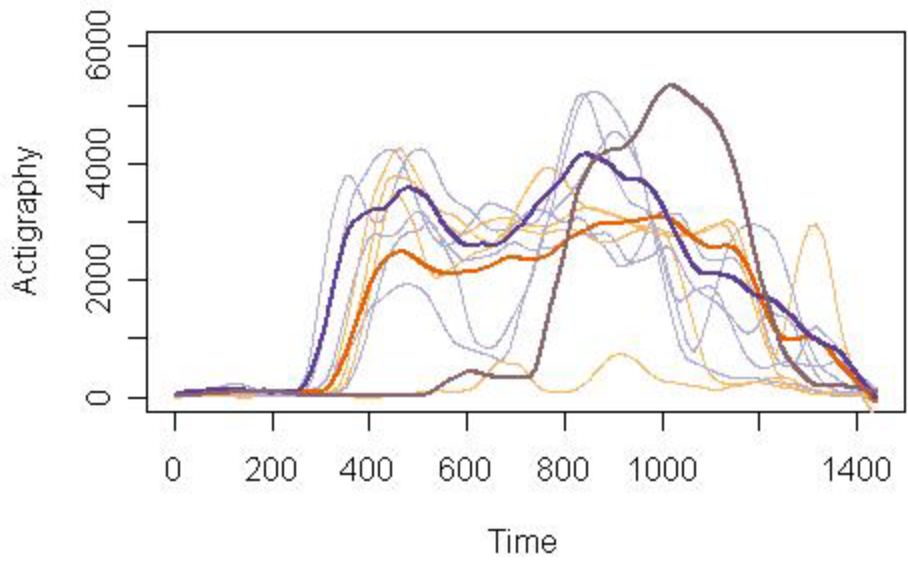
Velocity (First Derivative) of Smoothed Daily Data (Base Day 3 Brushed)



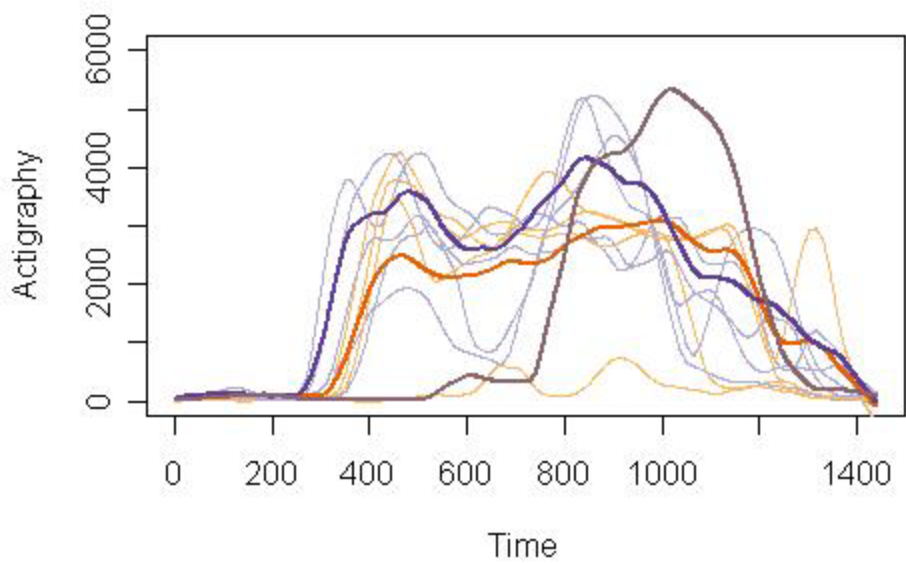
Acceleration (Second Derivative) of Smoothed Daily Data (Base Day 3 Brushed)



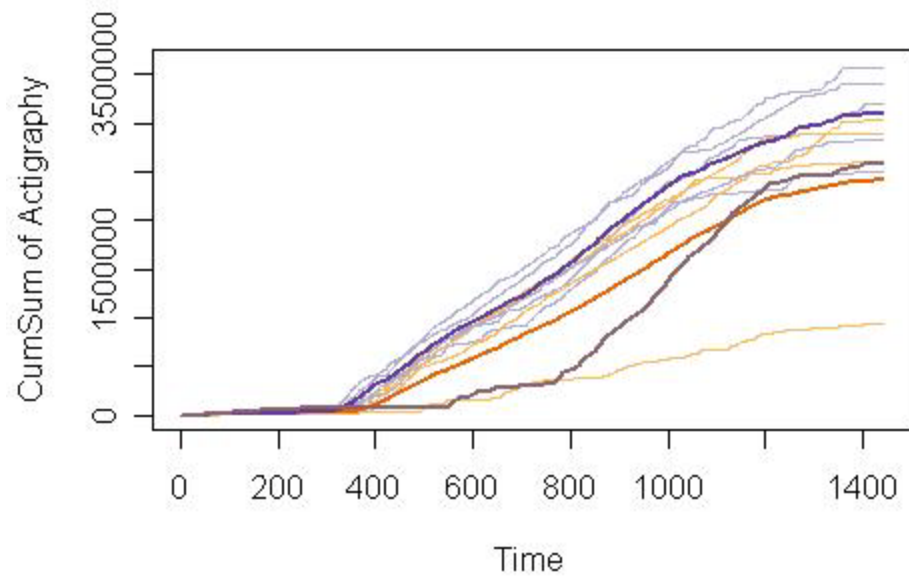
Smoothed Daily Data (Base Day 2 Brushed)



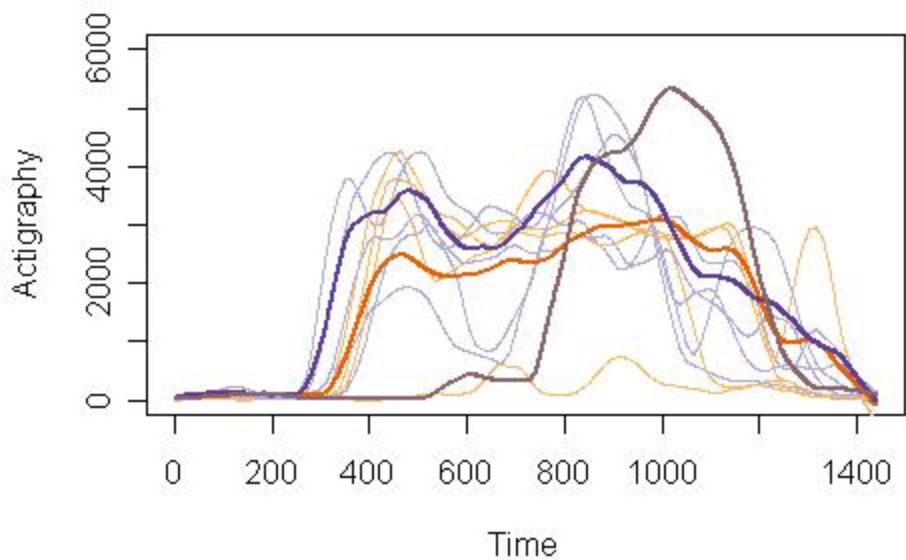
Smoothed Daily Data (Base Day 2 Brushed)



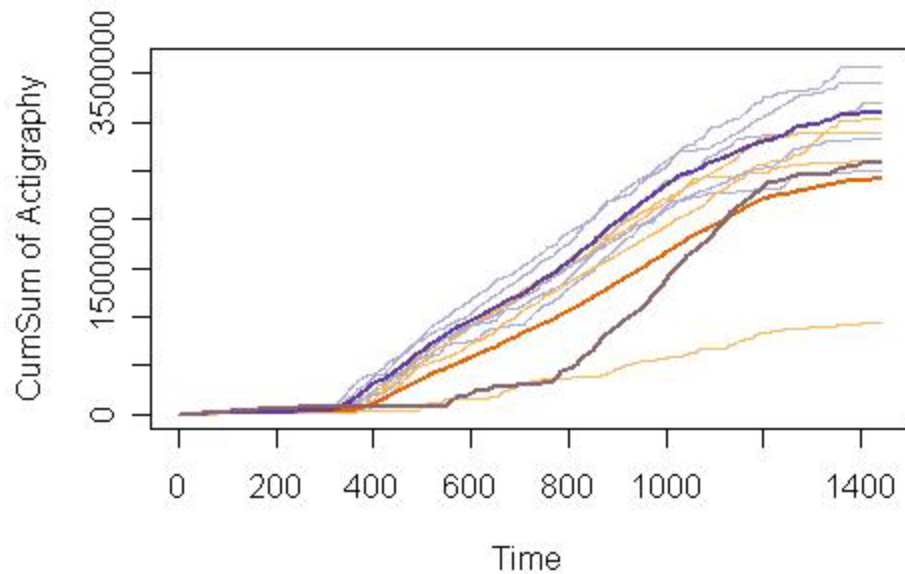
Cumulative Sums (Base Day 2 Brushed)



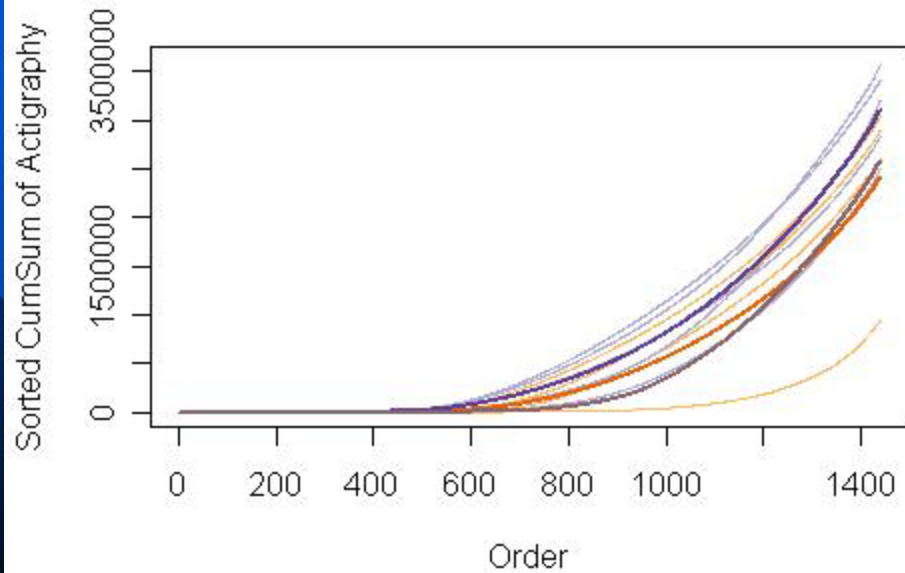
Smoothed Daily Data (Base Day 2 Brushed)



Cumulative Sums (Base Day 2 Brushed)



Sorted Cumulative Sums (Base Day 2 Brushed)



Conclusions

Visualization of Actigraphy Data provides

- Potential for application in various medical fields
- Additional insights into actigraphy data
- Ease to compare baseline and past-treatment data
 - » of a single patient
 - » of multiple patients
 - » to identify outliers
 - » to compare averages

■ **Example:**

Agreement in Carpal Tunnel Syndrome (CTS) Studies

Reference:

Dale, A.M., Strickland, J., Symanzik, J., Franzblau, A., Evanoff, B. (2008): Reliability of Hand Diagrams for the Epidemiologic Case Definition of Carpal Tunnel Syndrome, Journal of Occupational Rehabilitation, 18(3):233-248.

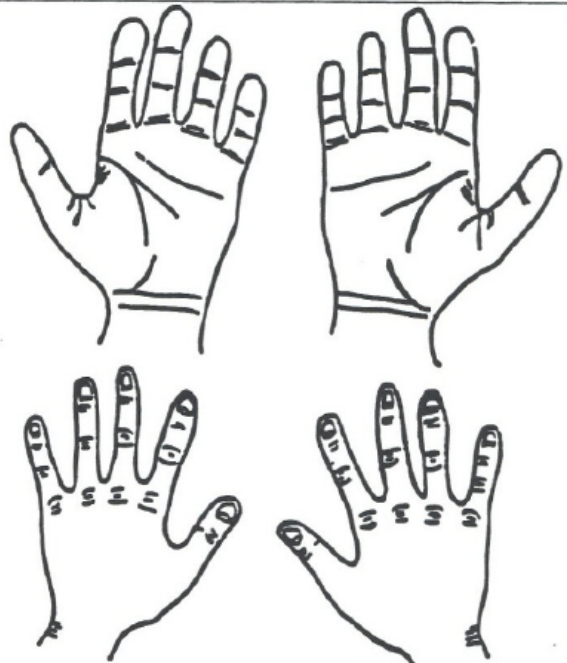
Study A Background

- Subjects have to indicate where they experience what kind of “pain” on their fingers/hands
- Background:
 - 393 Subjects
 - 494 Hand Diagrams
- Expert raters assign values from 0 (unlikely) to 3 (classic), describing severity of CTS
- Interested in rating agreement among 3 expert raters

	RIGHT WRIST	LEFT WRIST	RIGHT HAND	LEFT HAND	RIGHT FINGERS	LEFT FINGERS
Burning/ Pain	○	○	○	○	○	○
Tightness/ Stiffness	○	○	○	○	○	○
Soreness/ Cramping/ Aching	○	○	○	○	○	○
Numbness/ Tingling	○	○	○	○	○	○

Please show on the diagram to the right where you have experienced numbness, tingling, burning, or pain by shading in the problem area.

If you have not experienced these symptoms, please skip to the next question.



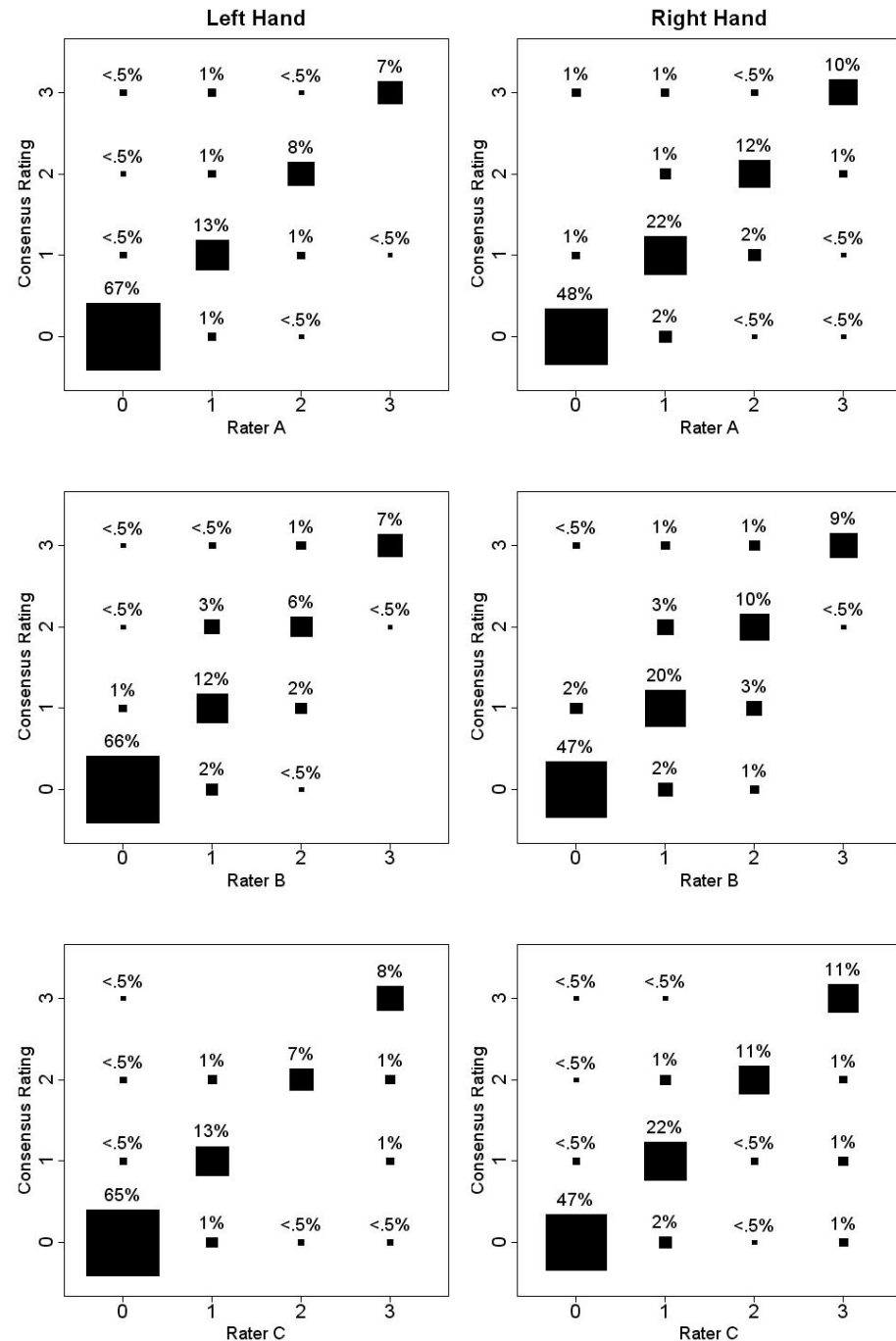
The diagram consists of four line drawings of hands. The top row shows the palms of the right and left hands. The bottom row shows the backs of the right and left hands. Each hand has a wrist line and a series of lines representing the fingers. The fingers are labeled with letters: T (Thumb), I (Index), M (Middle), R (Ring), and P (Pinky).

Numerical Results (A1)

- Weighted “kappa” represents agreement between three experts, based on difference between how much agreement is present compared to how much agreement would be expected by chance alone (possible range is -1.0 to 1.0)
- Interpretation:
 - < 0 : Less than chance agreement
 - 0.01 – 0.20: Slight agreement
 - 0.21 – 0.40: Fair agreement
 - 0.41 – 0.60: Moderate agreement
 - 0.61 – 0.80: Substantial agreement
 - 0.81 – 0.99: Almost perfect agreement
- Observed Results:
 - Left Hand Diagrams: 0.88 (95% CI: 0.83, 0.91)
 - Right Hand Diagrams: 0.83 (95% CI: 0.78, 0.87)

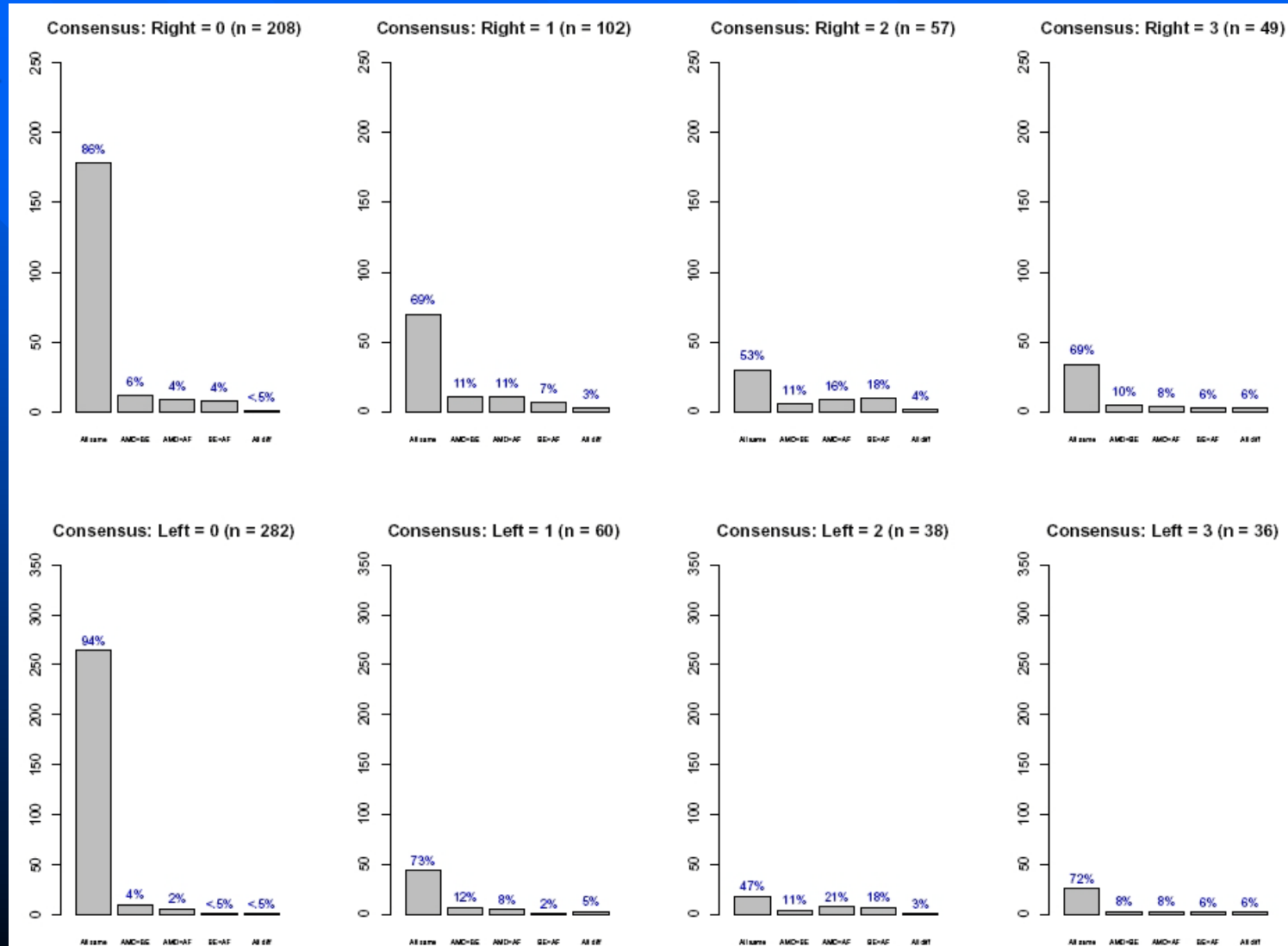
Graphical Results (A1)

- Comparison of Rating Consensus Scores of 3 Raters, Compared to the Raters' Individual Scores



Graphical Results (A2)

Detailed Comparison of Raters' Individual Scores



Study B Background

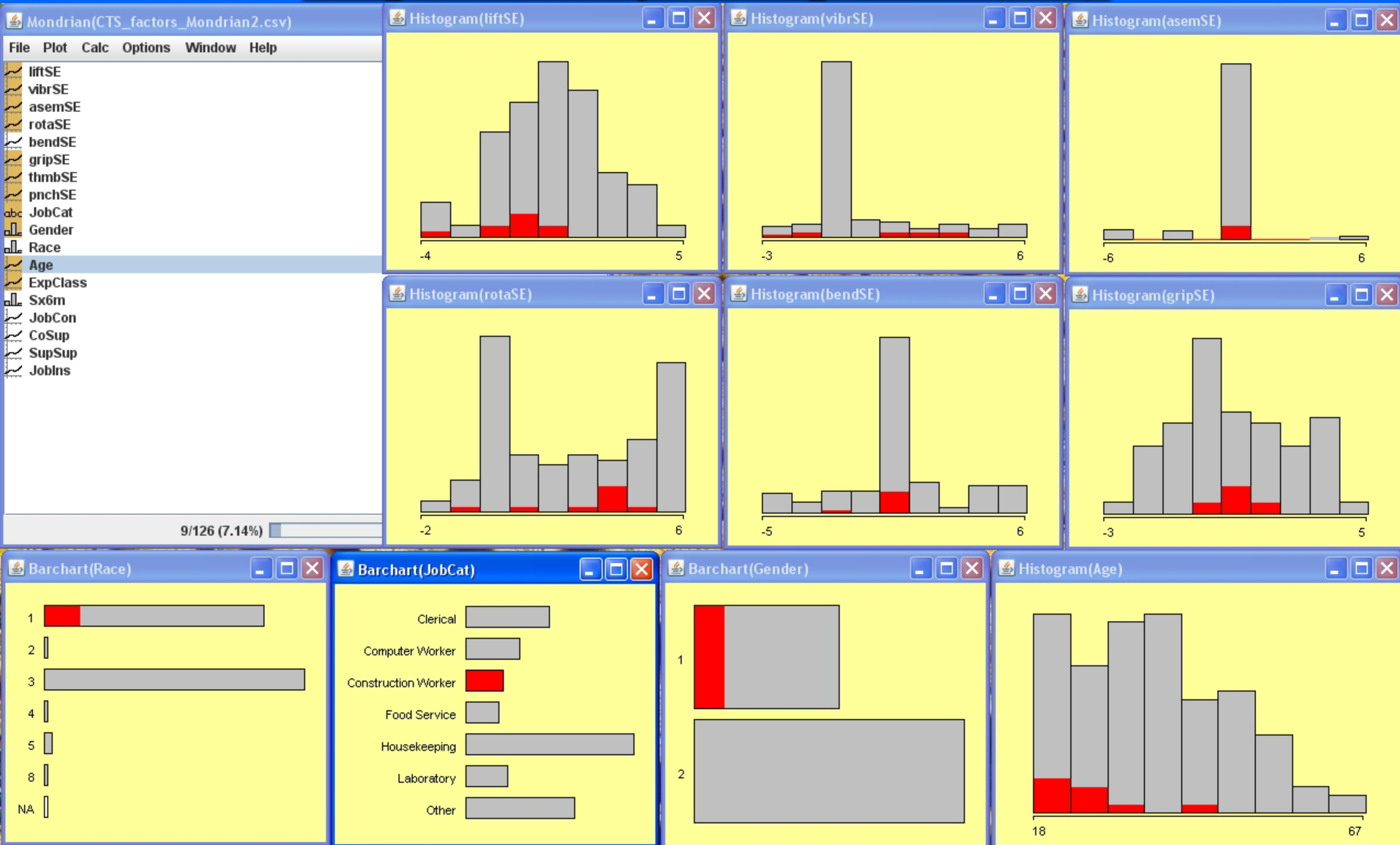
- Subjects have to indicate how much time they spend on typical tasks during a work day
- Subjects are videotaped and expert raters estimate the times the subjects spent on those tasks based on the videos
- Subjects and expert raters assign values from 1 (not at all) to 7 (> 4 hours)
- Interested in rating agreement among subjects and expert raters

Numerical Results (B)

Table 5. Weighted kappa coefficients for comparison of self-reported and observer-rated exposures for 341 subjects

Physical Exposures	Kappa coefficient	95% CI
Lifting more than 2 pounds	0.67	0.60, 0.73
Hand-held vibrating power tools	0.61	0.54, 0.68
Forceful hand grip	0.58	0.51, 0.64
Bending of hand/wrist	0.23	0.11, 0.34
Finger pinch grip	0.16	0.08, 0.24
Forearm rotation	0.04	0.003, 0.08
Assembly line tasks	-0.01	-0.08, 0.11
Finger or thumb pushing or pressing	-0.07	-0.18, 0.04

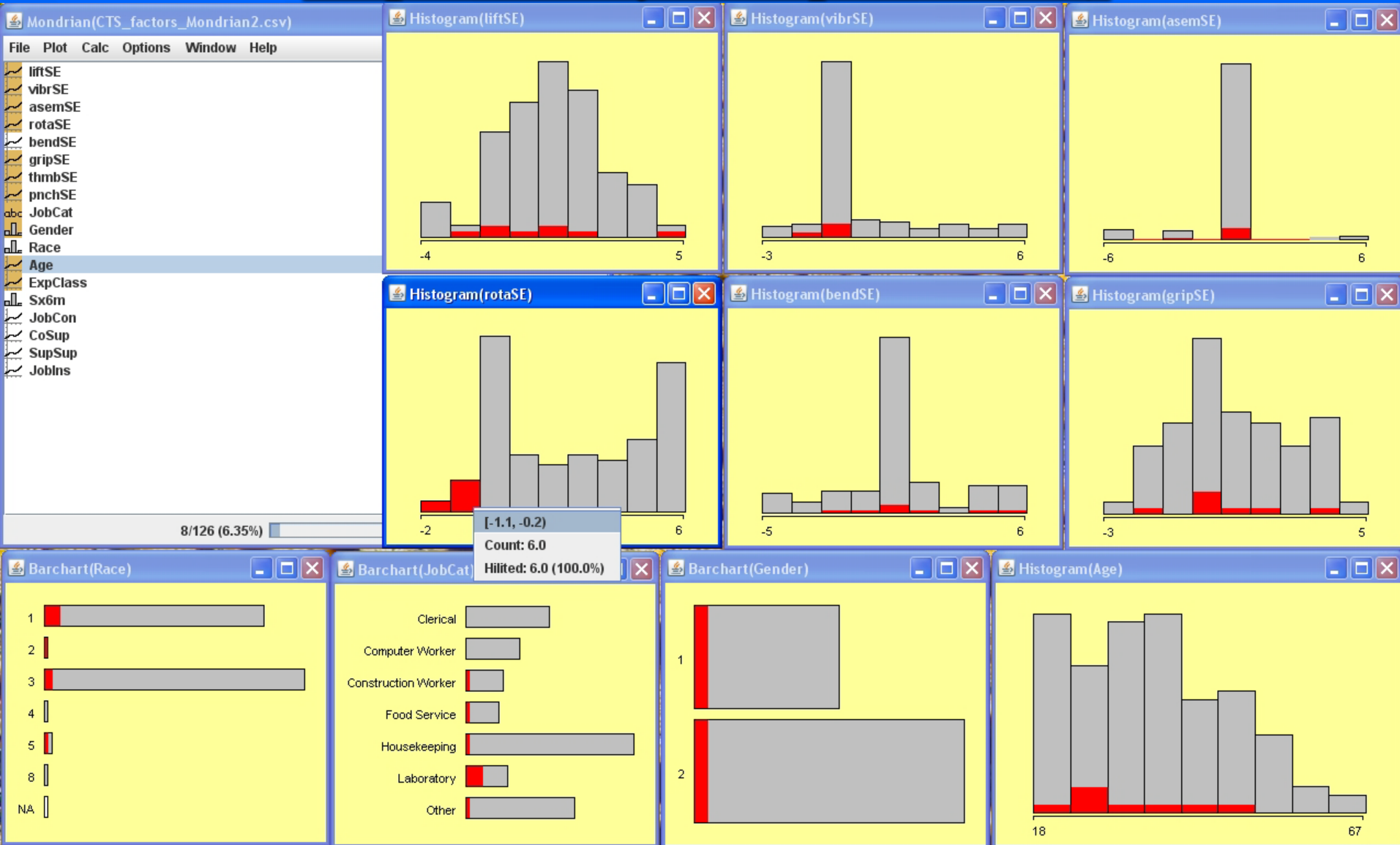
Exploratory Graphics (B1)



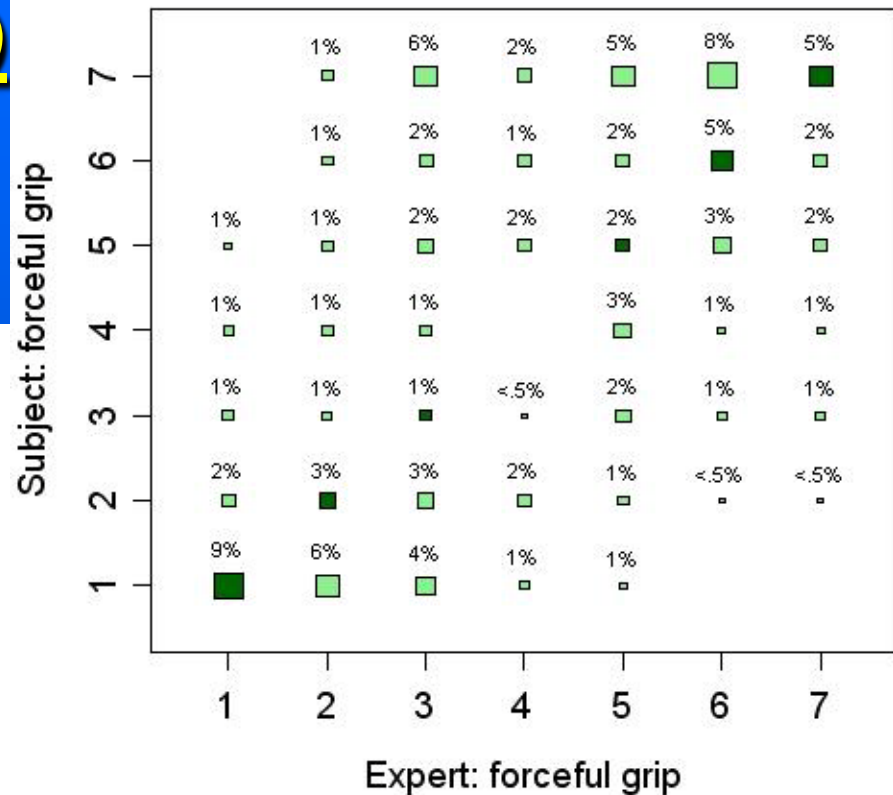
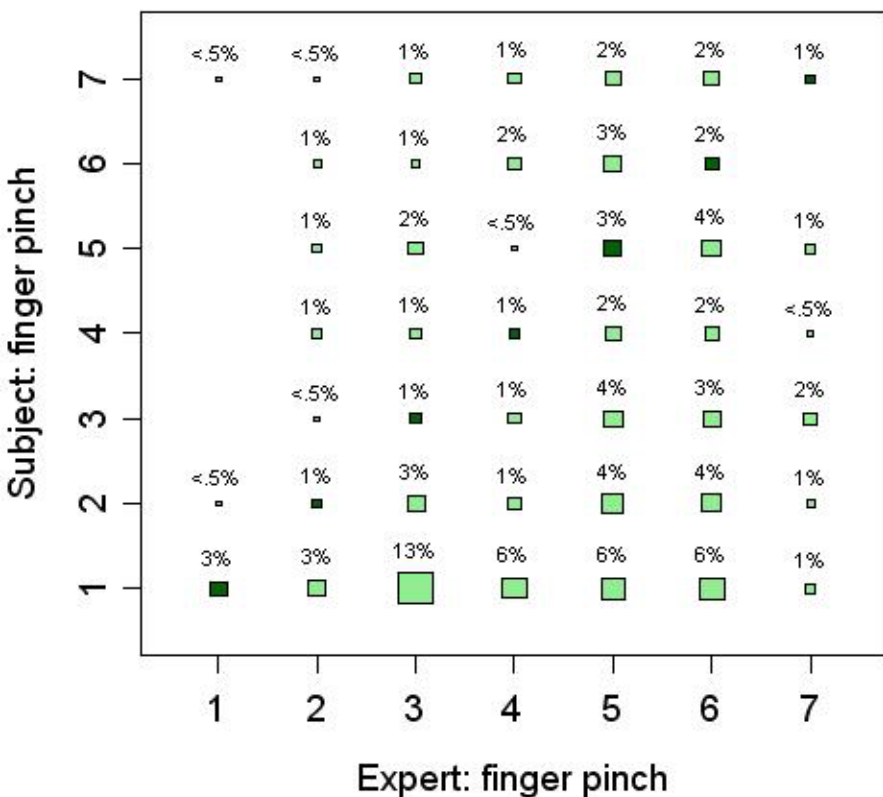
Exploratory Graphics (B2)



Exploratory Graphics (B3)

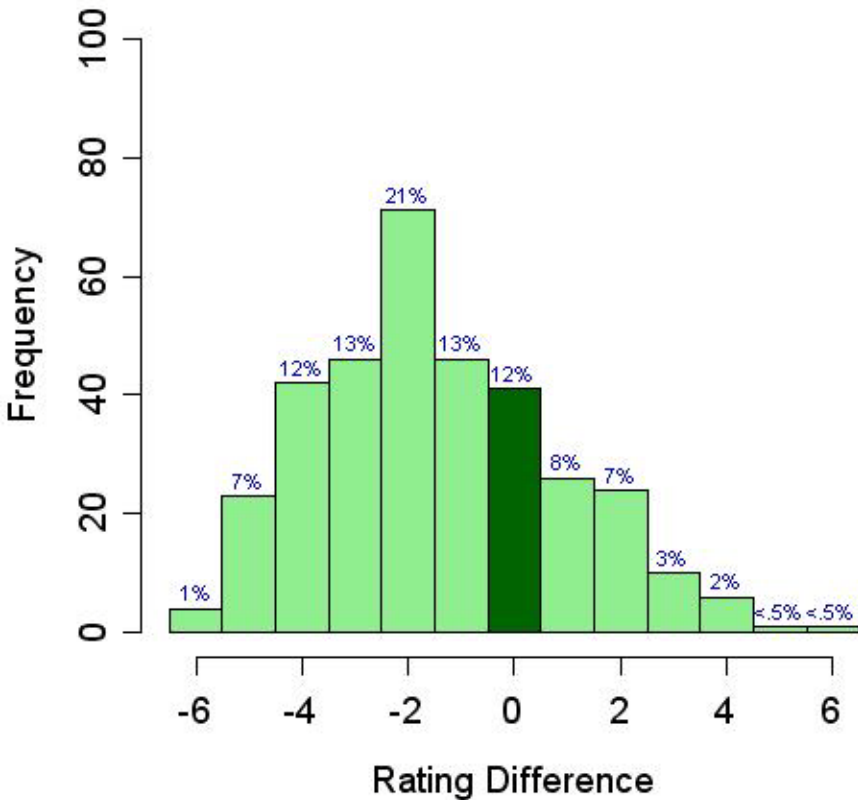


Graphical Results (B1)

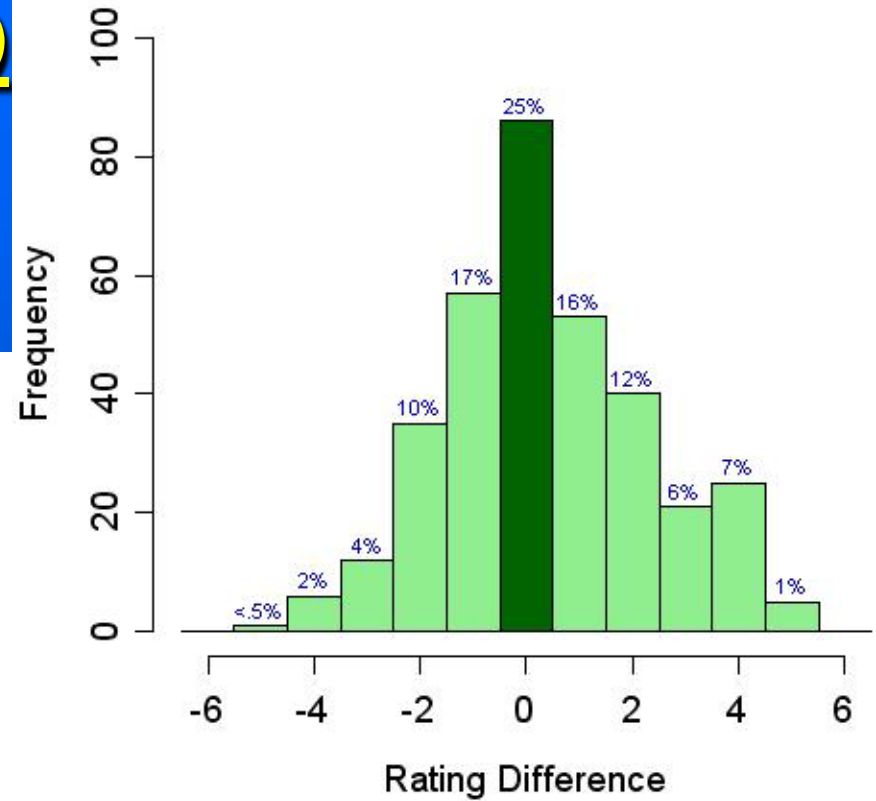


Graphical Results (B2)

Finger pinch: Subject - Expert [All Jobs]

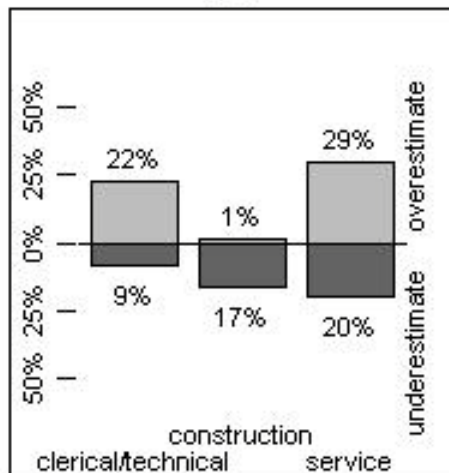


Forceful grip: Subject - Expert [All Jobs]

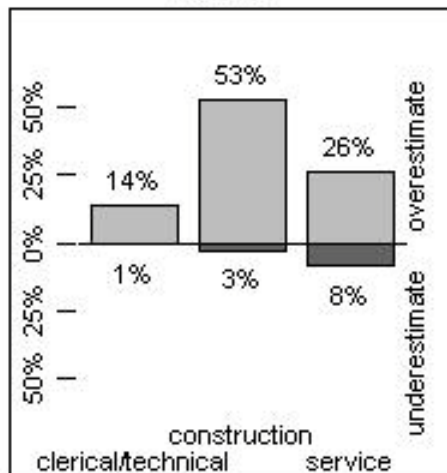


Graphical Results (B3)

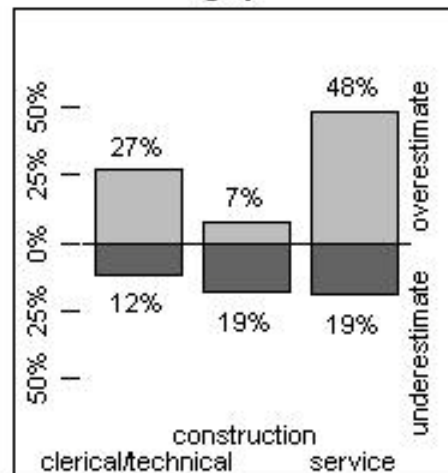
lift



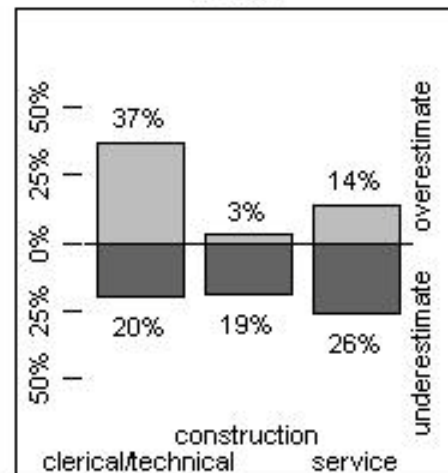
vibrate



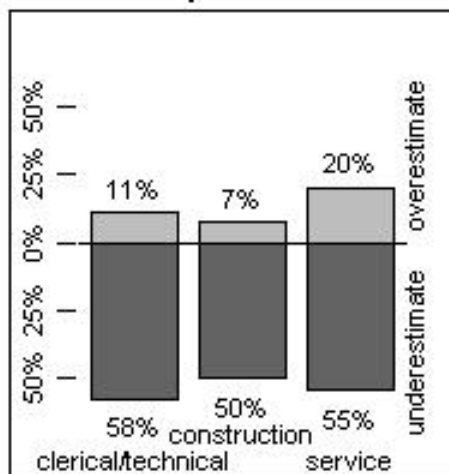
grip



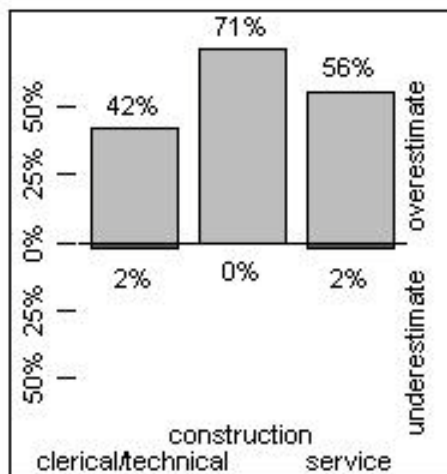
bend



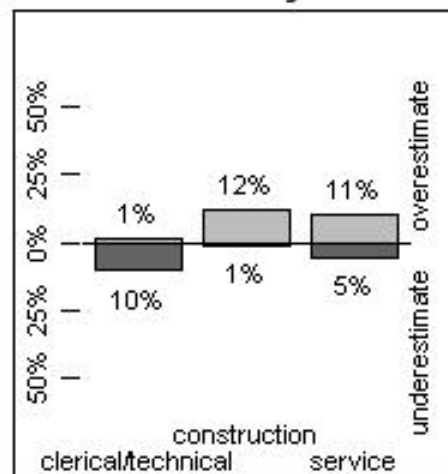
pinch



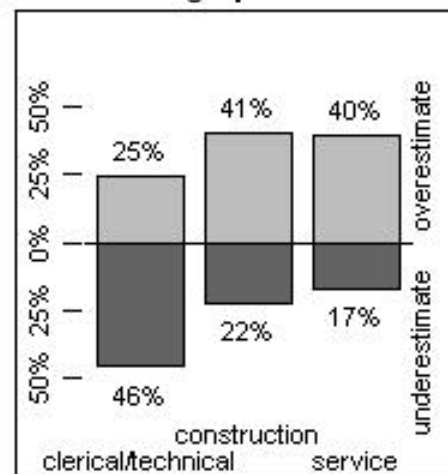
rotate



assembly



digit press



- **Example:**

Graphics for cgh and Gene Expression Arrays

- **Reference:**

Symanzik, J., Shannon, W. (2007): How Graphics can be Useful for the Simultaneous Exploration of cgh Array and Gene Expression Data, Talk presented at the Interface 2007 Conference in Philadelphia, PA (June 2007).

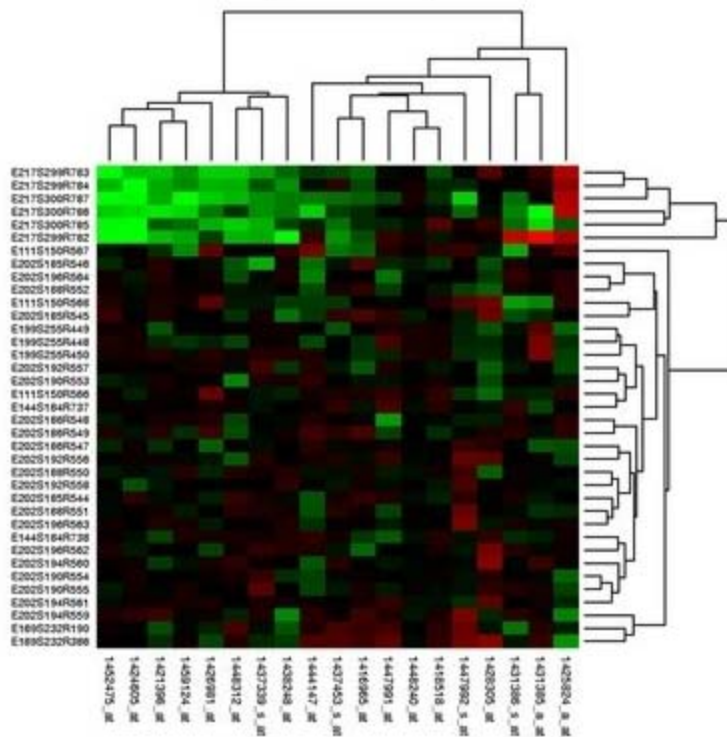
Side Note: Heatmaps for People with Color Vision Deficiencies

- From <http://www.vischeck.com/vischeck/vischeckImage.php>

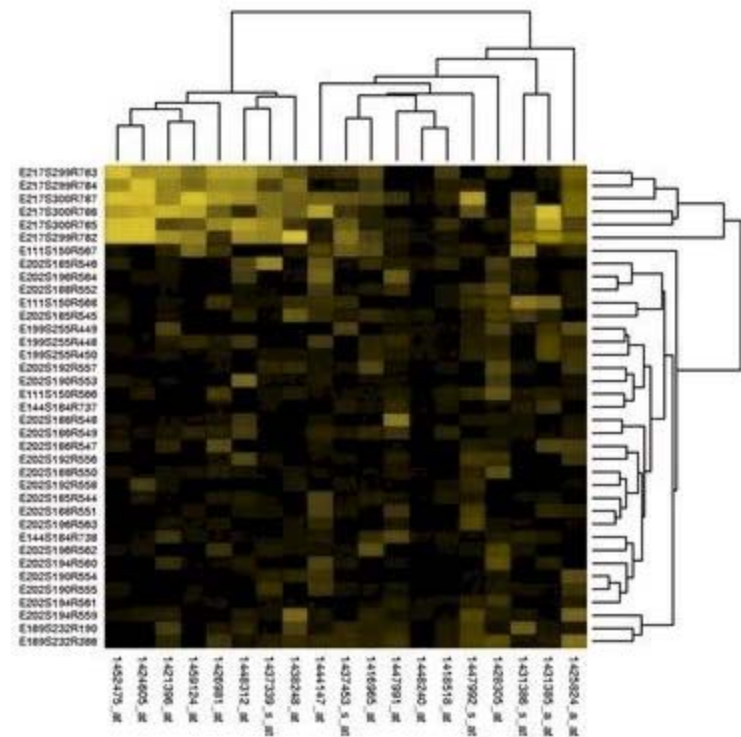
Try Vischeck on Your Image Files

Your Results:

Original Image



Deuteranope Simulation



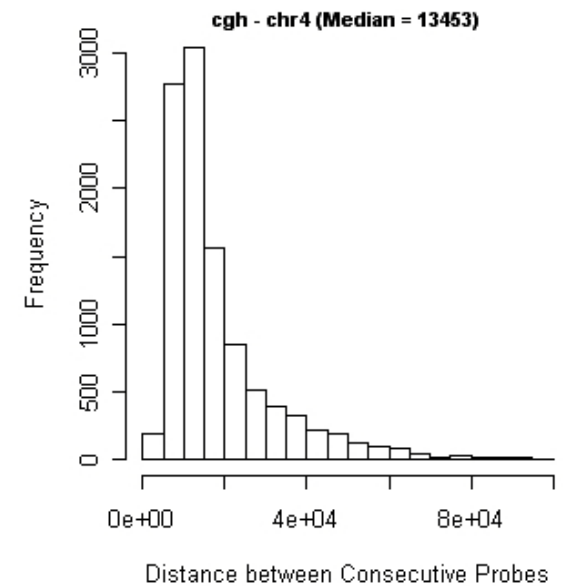
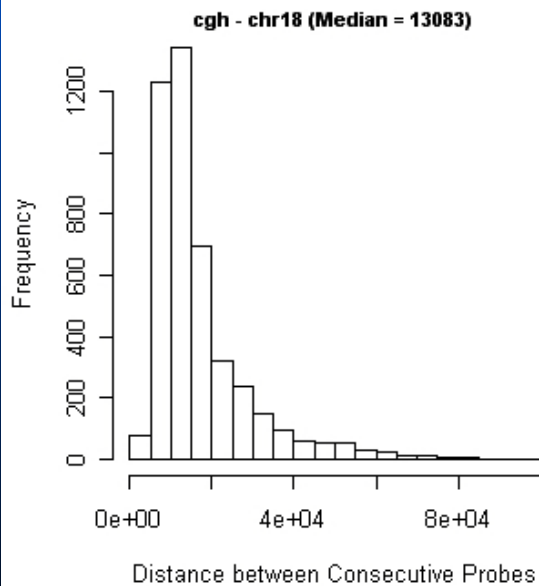
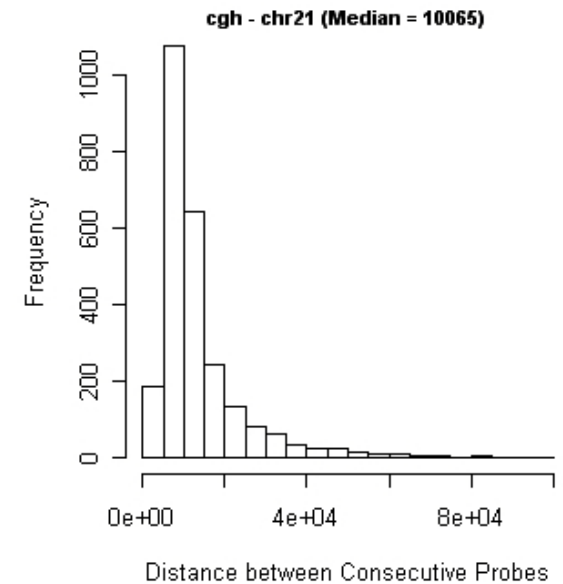
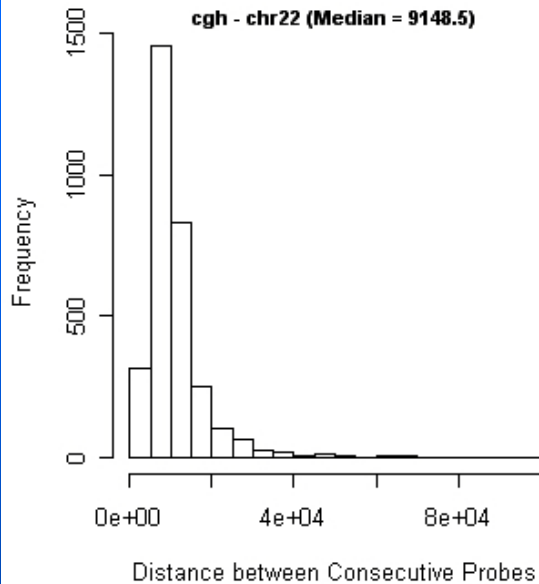
Background

- Data from 21 cancer cell lines
 - cgh array data:
 - Agilent, 181,984 probes per sample
 - Gene expression data:
 - 40,511 probes per sample
- Data and helpful discussion provided by Matthew Ellis and his lab staff, Washington University School of Medicine, St. Louis, MO

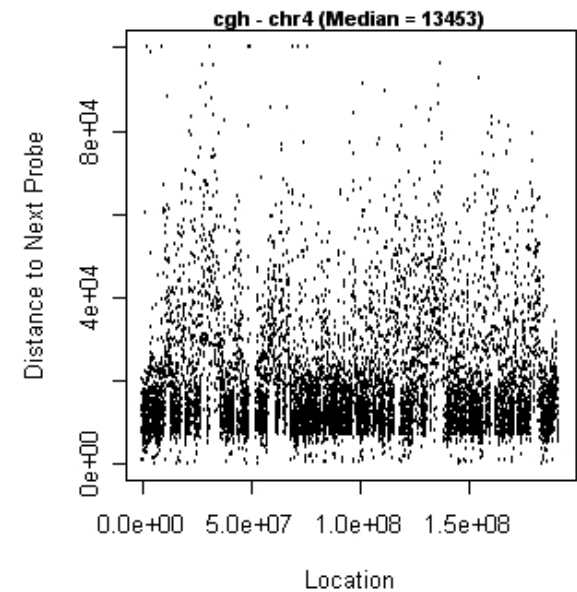
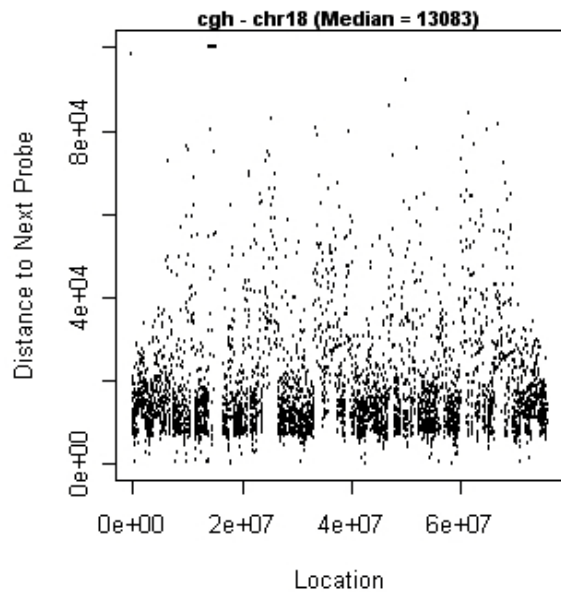
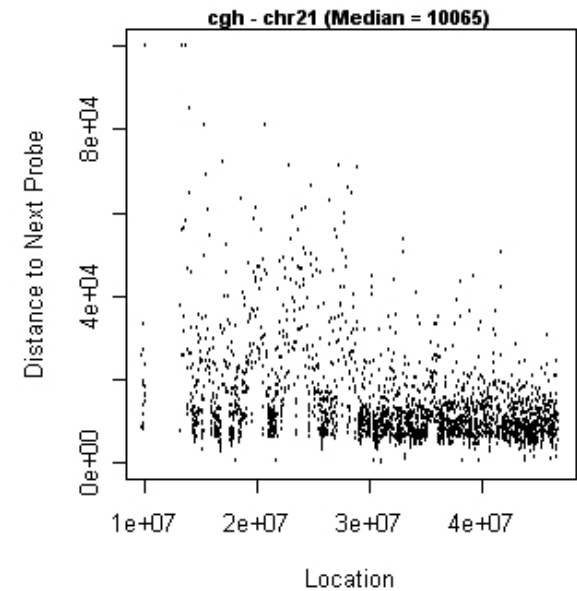
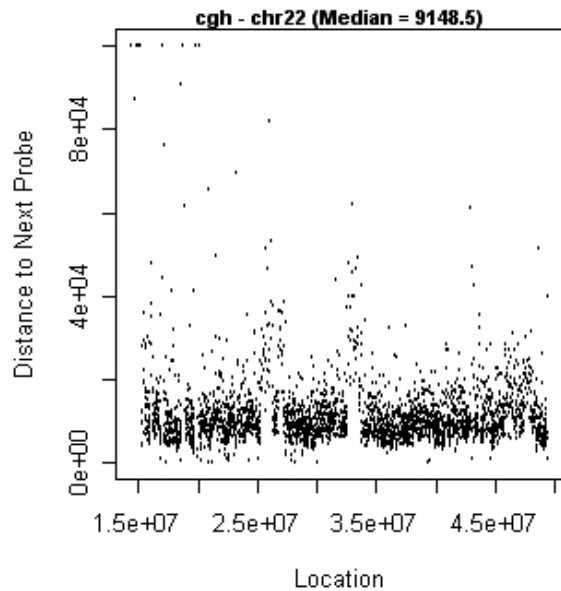
Question of Interest

- Can we identify regions on the chromosomes where high (low) values of cgh array are associated with high (low) values of gene expression data?
- Note:
 - Only about 1/5 of gene expression probes
 - cgh array and gene expression probes not at exactly the same locations

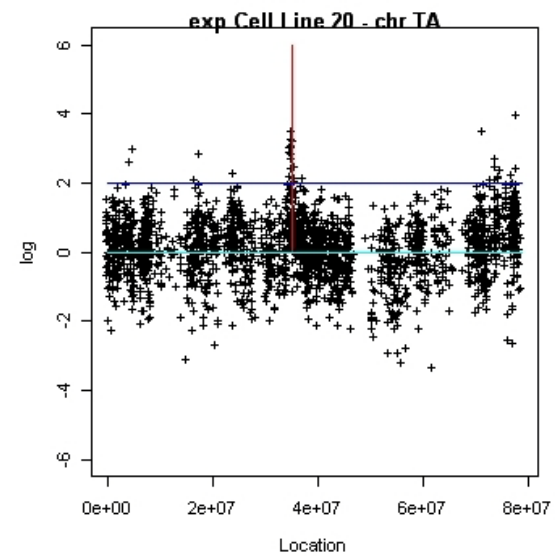
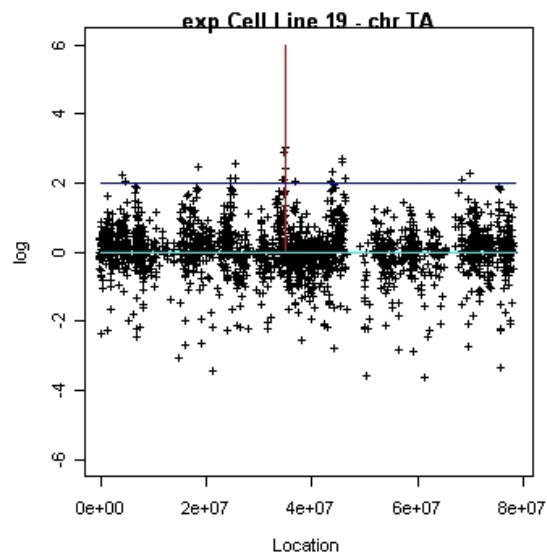
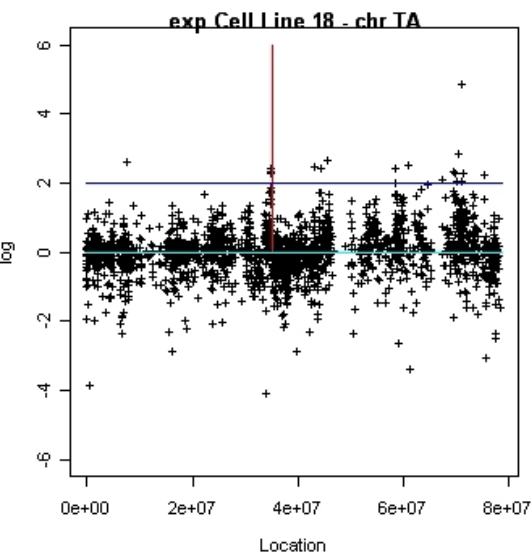
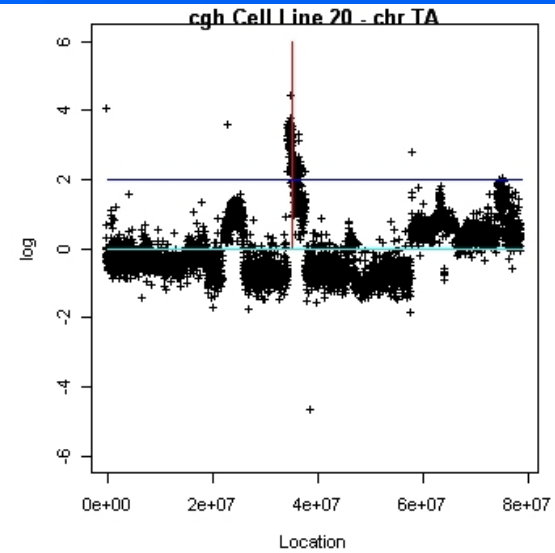
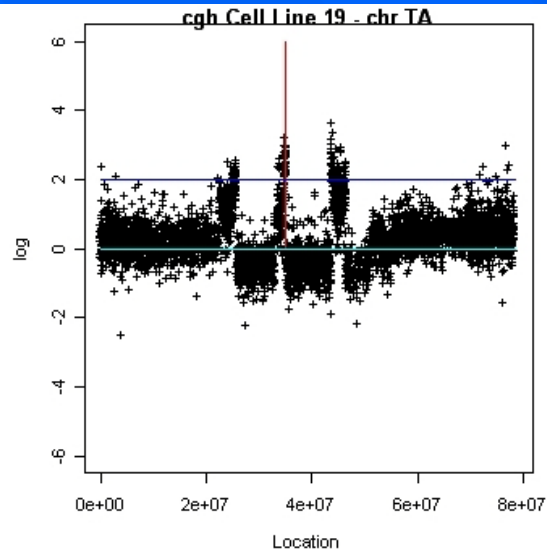
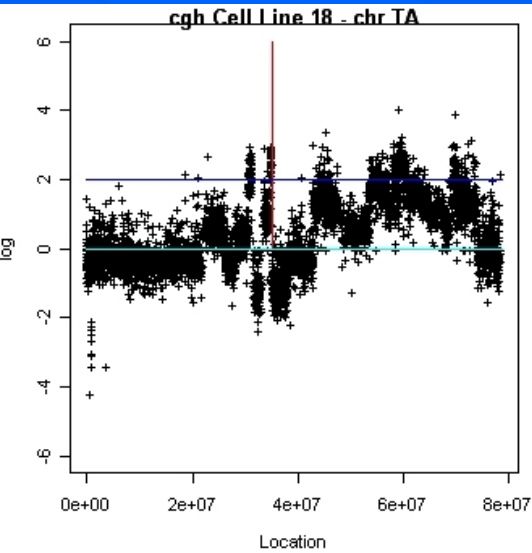
cgh Probe Spacing



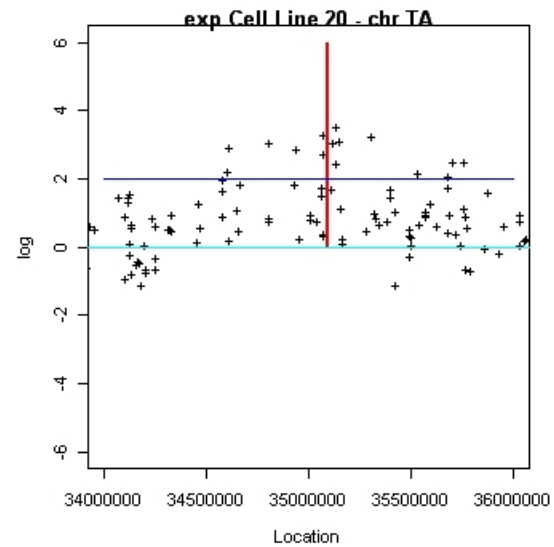
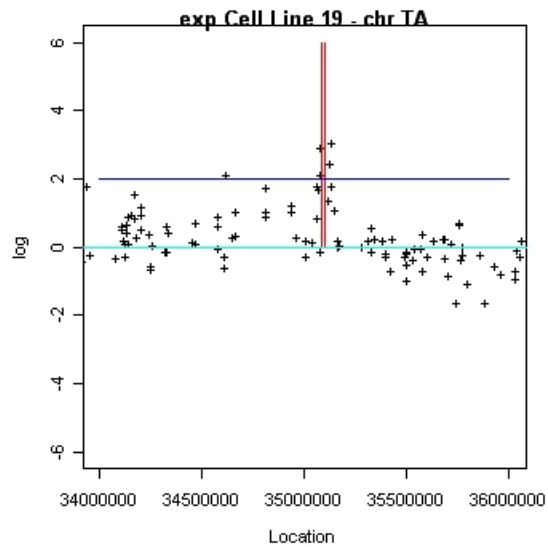
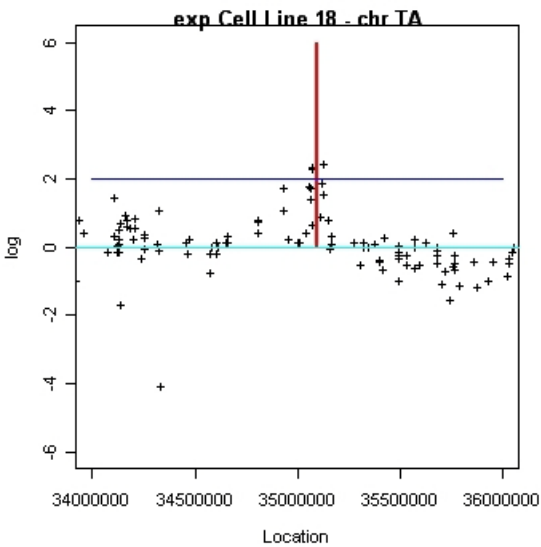
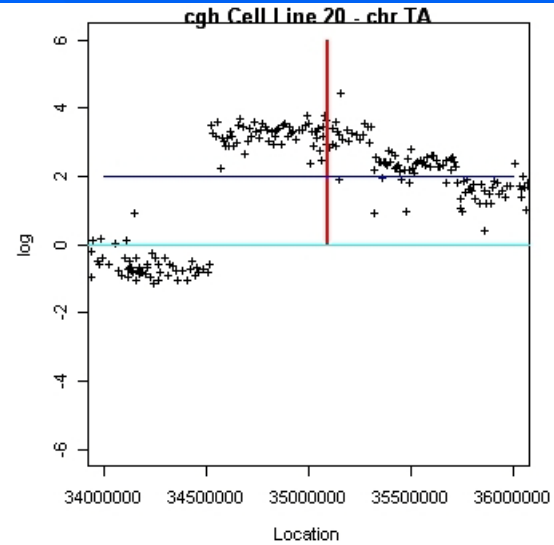
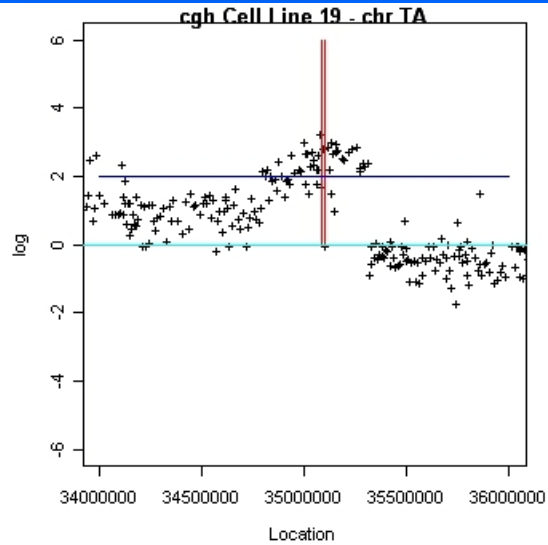
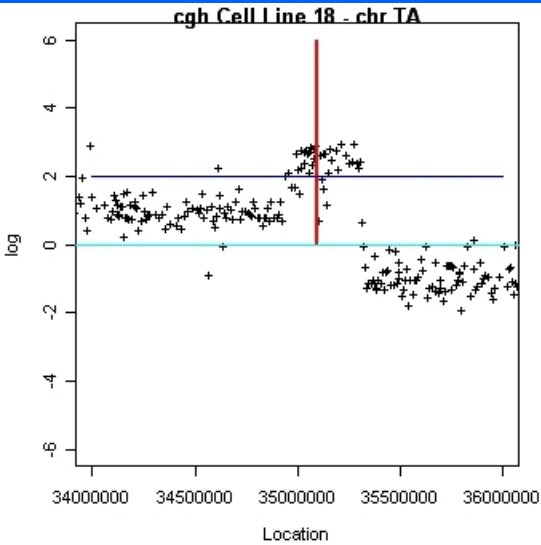
cgh Probe Spacing (2)



A Well-Known Region



A Well-Known Region (2)



A Well-Known Region (3)

- Observations:
 - 3/21 cell lines have unusually large cgh values (≥ 2.0) at 2 nearby locations on this chromosome
 - Gene expression values at closest location each also above a threshold (≥ 2.0)
 - Other cgh values / gene expression values in this region not necessarily above these thresholds

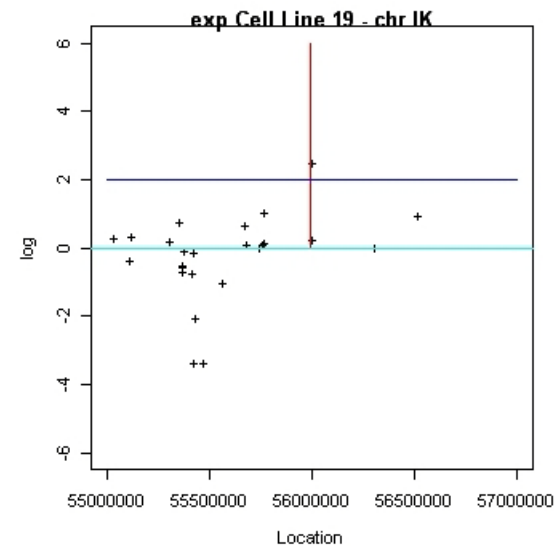
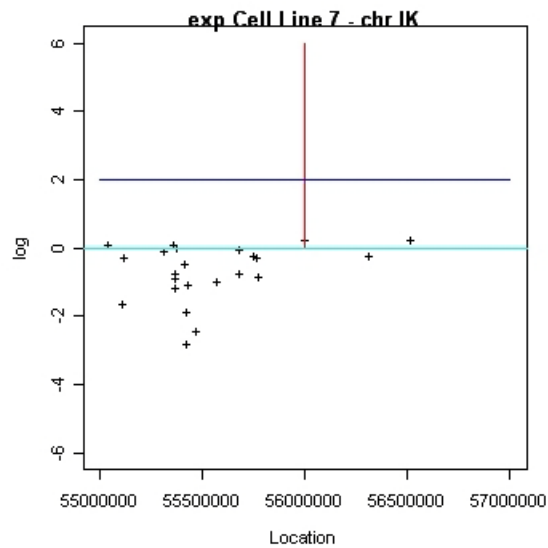
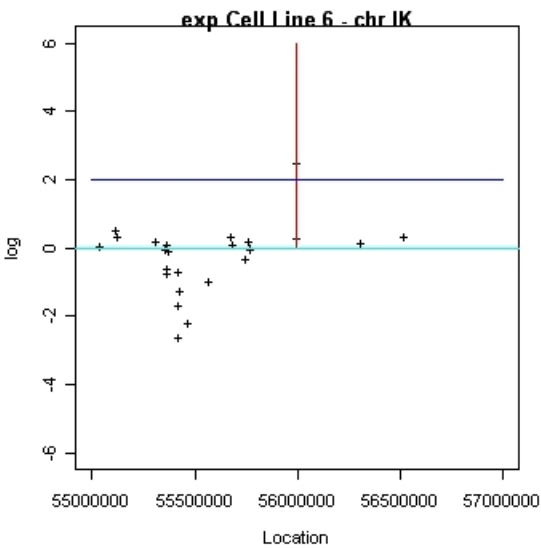
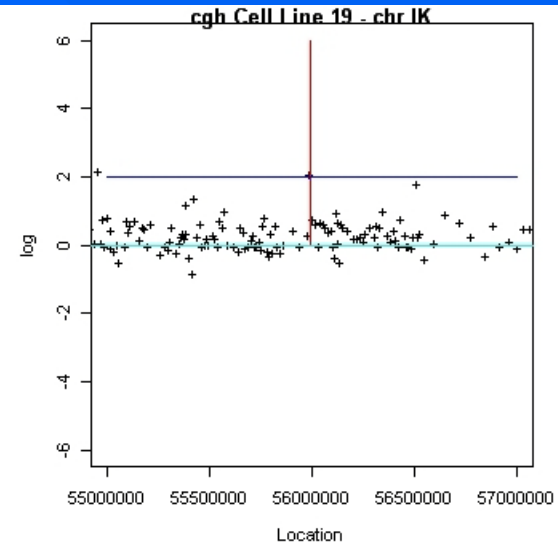
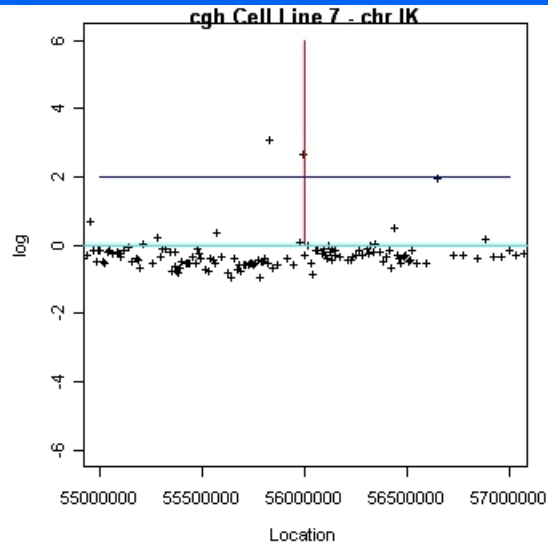
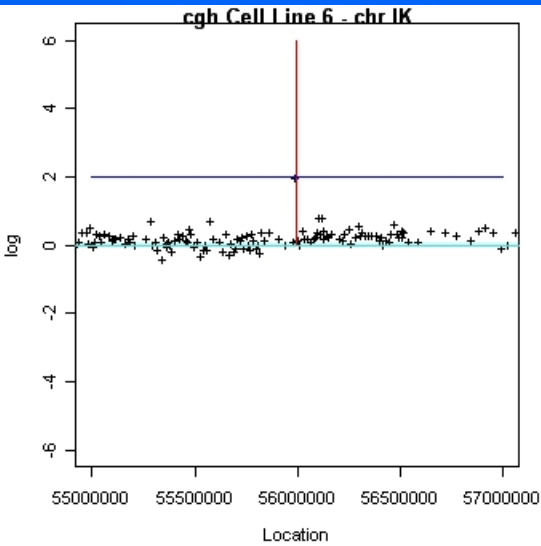
Simultaneous Peak Search

- Motivation:
 - Most approaches discard single cgh and/or gene expression peaks as noise
 - What if no noise, but observed on several cell lines?
 - Keep in mind: unequal spacing between probes!

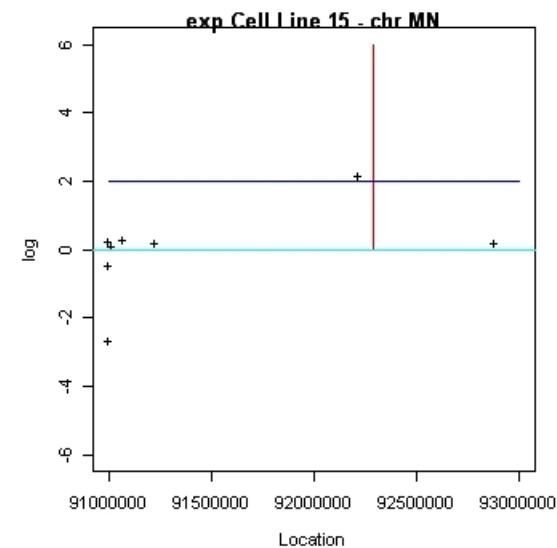
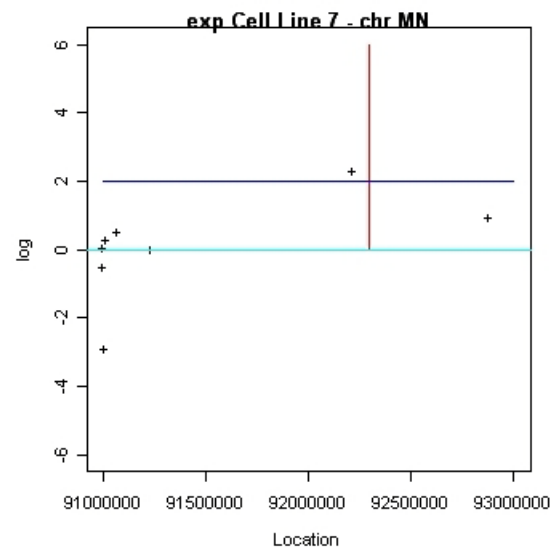
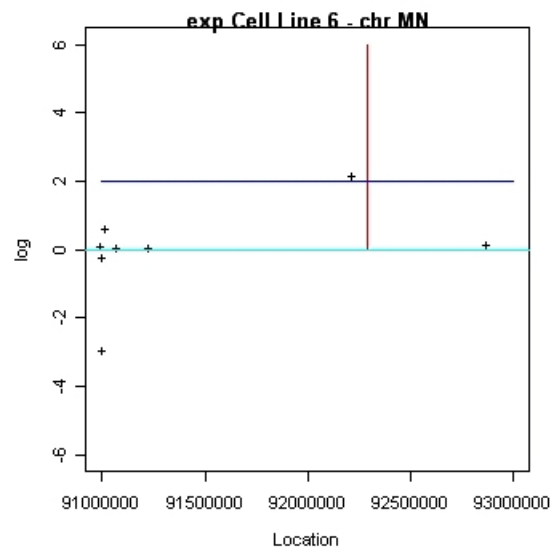
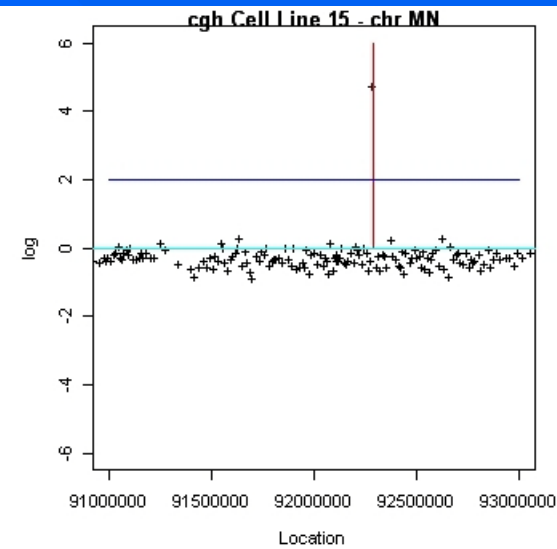
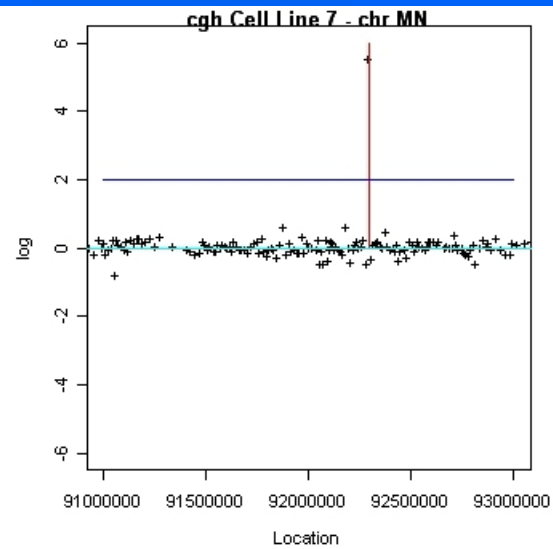
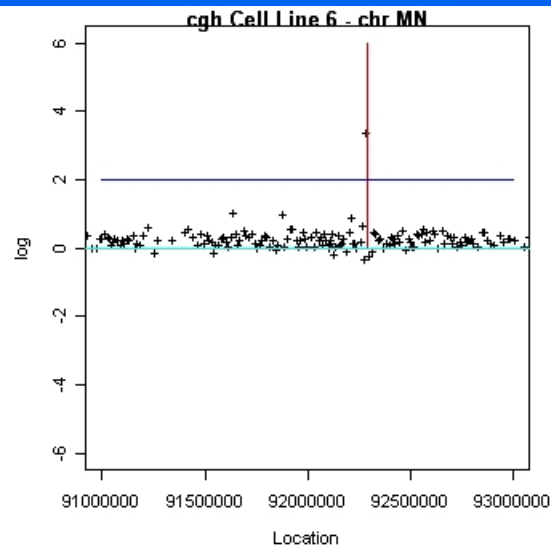
Simultaneous Peak Search (2)

- Approach:
 - For each cgh location, determine nearest exp location; consider pairs (cgh value at cgh location, exp value at nearest exp location)
 - Consider cgh/exp value pair as unusual (=success) iff cgh value ≥ 2.0 & exp value ≥ 2.0
 - List all successes; sort by number of affected cell lines

Successes



Successes (2)



Overall Conclusions

- Visual approach effective to see unexpected structure in data
- Not every graphic will make it into a paper, but many of those that do not make it will provide useful insights for the researchers:
 - Graphics often help to create new hypotheses
 - Graphics useful to verify results
- Combination of different graphical techniques (static, interactive, various plots) most effective
- Graphics can be used for almost all types of data

Questions ???