



DATA SCIENCE & DATA VISUALIZATION

Copyright © by @itdsiu20031. All rights reserved. Do not redistribute without consent.
Nguyen Quang Dieu, Nguyen Hoang Anh Tu
Jan 16, 2024

Table of contents

01 Theory (15 mins)

03 Code (60 mins)

02 Visualization (30 mins)

04 Test bank (30 mins)





01

THEORY



Copyrighted by @itzmealvin - ITDSIU20031.
Do not be distributed without consent

Critical points of W07 - Interaction

- 
1. Static content: infographics, books + dynamic content: animated in auto-play, interactive content
 2. Need to interact because exploring data that is big/complex -> amplifies cognition
 3. Direct manipulation: interact directly with objects + indirect interact
 4. Types: single view: overtime, navigation, semantic zooming, filtering, focus + multiple views: selection, linking, brushing, adapting representation
 5. Change over time: use slides to see views at different times, show differences explicitly -> doesn't have to be literal time
 6. Transition: change orders, track what's going on -> animated transitions: smooth interpolation between states/techniques
 7. Animation caveats: changes hard to track, and eyes over memory
 8. Navigation: pan, zoom, rotate
 9. Scrolltelling: an interactive story, interacting by scrolling but unexpected behavior
 10. Semantic zooming: update content on zooming, more details and readable at any resolutions
- 

Critical points of W07 – Interaction

- ✚ 11. Focus + Context: pick what to show, hint not showing -> visual encoding and interaction (aggregation, reduction, layering)
- 12. Elision: focus item shown in detail
- 13. Degree of interest (DOI): Represent objects in the neighbourhood in detail and only major landmarks far away $DOI(x) = I(x) - D(x,y)$
- 14. Interactive tree with animated transitions that fit within a bounded region of space layout depends on the user's estimated DOI tree
- 15. Superimpose: Focus layer limited to a local region of view
- 16. Magic lenses: details data is shown when moving len over a scene -> Labeling
- 17. Distortion: Use geometric distortion of the contextual regions to make room for the details in the focus regions(s)
- 18. Distortion kinds: perspective wall, fisheye, hyperbolic geometry -> unsuitable for relative spatial judgments
- 19. Overview and detail: One view shows an overview + Other shows a detail
- 20. Filter & Dynamic querying: Mantra overview first, zoom & filter, details on demand



Critical points of W08 - Views

1. Multiple views: eyes beat memory, no single visual encoding is optimal, and too many to be shown in one view
2. Linked views: Multiple views are simultaneously visible and linked together (highlighting + navigation + encoding + dataset)
3. Multiform: Different visual encodings are used between views, supporting different tasks
4. Stack zooming, Mizbee, Stratomex -> Small multiples: same visual encoding, but shows a different subset
5. Partitioning: Action on the dataset that separates the categorical data into groups (divide data + splits + views)
6. Trellis plots: panel variable (encoded in individual views), partitioning variables (assigned to columns and rows), main-effects ordering
7. Recursive subdivision: Flexibly transform data attributes into a hierarchy using treemaps as space-filling rectangular layouts
8. Layering/Overlay: combine multiple views on top of one another -> composite view
9. SUPERIMPOSED (best for tasks carried out within a local visual span) VS JUXTAPOSED (best for global tasks)
10. Dual axis, Combined chart, Layers/Dynamic layers





Critical points of W09 - Table

1. Table: scale (1000+ need analysis), records (10 000 need analysis), homogeneity (same types/scale)
2. Analytic component: scatter plot, parallel coordinated -> heat map -> multi-dimensional scaling
3. Techniques: magnitude (size comparison), distribution (aggregating large data), ranking (magnitude ranking, bump charts, temporal, table lens, lineup), part to whole, deviation (reference point), correlation, change over time (line chart, stacked area, sparklines, clipped graphs, heatmap)
4. Bar chart and isotype visualization -> Part of the whole: Show how a single entity can be broken down into its component elements, stacked bar chart, pie and donut chart, treemap, stacked area
5. Histogram -> good choice of bins = \sqrt{n} or $\log_2(n) + 1$, density plot, box-and-whisker plot, notched box plot (with confidence interval), dot plots, violin plot = box plot + probability density function
6. Multiple attributes: combiner function (weighted sum for serial, maximum for parallel, and product/nesting for complex)



Critical points of W10 - Storytelling

- 
1. Good stories = facts + data + context + engage + educate
 2. Underscore your arguments with Data/Facts and leverage the power of Visualization
 3. Components: Introduction, Context, Main story, Annotation of key point
 4. Genre: magazine, annotated chart, partitioned poster, flow chart, comic strip, slide show, film/video/animation
 5. Author driven << linear ordering, heavy message, no interactivity + no ordering, no message, free interactivity >> Reader driven
 6. Martini glas (author-driven, then open to explore), interactive slideshow (multiple scenes, interaction midway), drill-down story (decide path, annotated)
 7. Layout: descriptive titles, subtitles, annotation, saturation
 8. Interactivity: navigation, details on demand, relevant to the reader -> ask for opinions/prior knowledge
 9. Design: fewer colors, average for context, better scale, richer annotations
 10. Engagement: know target audience (opinionated <-> high information density) -> public media, expert panel, education, group meeting, board meeting
- 

Critical points of W11 - Evaluation

- ✚ 1. Problem-driven: top-down approach, identify a problem encountered by users, design a solution to help users work more effectively sometimes called a design study
- 2. Technique-driven: bottom-up approach, invent new visualization techniques or algorithms, classify or compare against other idioms and algorithms
- 3. Nested model approach
- 4. Design process: domain problem -> map to task + data type & factors -> identify & implement suitable technique
- 5. Domain characterization: details domain, grouped users, target domain, questions, data
- 6. Domain problem: infinite domain task, broken down to abstract task -> solutions probably exist
- 7. Task abstraction: what-why, generalized terms, task that user wants to do, data types and model, transform data -> specific task requirements
- 8. Encoding & Interactions: design of visualization techniques, manipulation of visual representations, decisions of separated/intertwined, drive decisions



Critical points of W11 - Evaluation



9. Task: analyze, search, query

10. High-level Analyze: consume -> discover, present, enjoy & produce -> annotate, record, derive

11. Mid-level Search: target, location -> lookup, browse, locate, explore

12. Mid-level Query: one, some, all -> identify, compare, summarize

13. Low-level Target: all data (trend, outliers, features), attributes (one/many), network data (topology), spatial data (shape)

14. Design: creating something new to solve a problem; design is used in many fields

15. Function can constrain possible forms -> Form depends on tasks that must be achieved

16. When designing: wicked problems (no clear definition, not a good solution) / not-wicked problems (math. chess, puzzles)

17. Why matter? Ineffective visualization combinations, unique problems & data, tasks, design space



Critical points of W11 - Evaluation

18. Evaluation methods: controlled experiments, interviews/questionnaires, field/lab observations, log analysis, algorithmic performance measurement, heuristics evaluation, usability testing, Wizard of Oz, eye-tracker, expert, insights-bases, case studies

19. Quantitative: metrics, measurements, number/stats for data vs Qualitative method: subjective metrics, descriptions, understandings

20. Internal validity: can you trust your experiment (high when in lab conditions, affected by test conditions) vs External validity: representative of real-world usage (high when tested in the fields, valid in the world) -> trade-off: The more akin to real-world situations, the more experiment is susceptible to uncontrolled sources of variation

21. Scope of evaluation

22. Predesign -> user work environment and workflow, design -> visual encoding + interaction design, prototype -> see if it achieves design goals and compares with conventional solutions, deployment -> see if it affects workflow/work process and effectiveness in the fields, re-design -> improve current design by identifying usability problems



02

VISUALIZATION



Copyrighted by @itzmealvin - ITDSIU20031.
Do not distribute without consent

03

CODE

(open in VSCode)



Copyrighted by @itzmealvin - ITDSIU20031.
Do not redistribute without consent

04

TEST BANK



Copyrighted by @itzmealvin - ITDSIU20031.
Do not redistribute without consent

FEEDBACK FORM



Copyrighted by @itzm
Do not redistribute



Thank you for coming

DATA SCIENCE & DATA VISUALIZATION



Copyright © 2024 by @it247.com - All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage or retrieval system, without consent in writing from the publisher.

Nguyen Quang Dieu, Nguyen Hoang Anh Tu
Jan 16, 2024