Contents

0.1 Assessment . . . 1

1 Lecture Notes **3**

1.1 Data Visualization . . . 3
1.1.1 Visualisation Procedure . . . 3
1.1.2 Data acquisition . . . 3
1.2 Univariate data . . . 3
1.2.1 Descriptive Statistics . . . 3
1.3 Bivariate data . . . 3
1.4 Time-series . . . 3
1.4.1 Time series periodicity . . . 4
1.5 Colour, light, and animation . . . 4
1.5.1 Colour . . . 4
1.5.2 Animation . . . 4
1.6 3D and 4+D Visualisation . . . 4
1.7 Multivariate data analysis . . . 5
1.8 Spatial statistics . . . 5
1.8.1 Measures of dispersion . . . 5
1.8.2 Nearest neighbour analysis . . . 5
1.9 Multivariate data analysis . . . 5
1.9.1 Principal Component Analysis (PCA) . . . 5
1.9.2 Cluster Analysis . . . 5
1.10 Visualising multidimensional data . . . 5
1.10.1 Ternary plots . . . 5
1.10.2 Scatterplot matrix . . . 5
1.10.3 Coplots . . . 5
1.10.4 3d point plots . . . 5
1.11 Computer Graphics (CG) . . . 6
1.11.1 What is CG . . . 6
1.11.2 Offline . . . 6
1.11.3 Interactive . . . 6
1.11.4 Real Time . . . 6
1.11.5 VR / AR . . . 6
1.12 Coordinate Systems . . . 6
1.12.1 Euclidean Space . . . 6
1.12.2 Points and Vectors . . . 6
1.13 Transformations . . . 6
1.13.1 Translation . . . 6
1.13.2 Rotation . . . 6
1.13.3 Scale . . . 7
1.13.4 Handedness . . . 7
1.14 Spaces . . . 7

Contributors:

- Daniel Fitz (Sanchez)

0.1 Assessment

- Two Projects
  - Visualization
  - Graphics
- Each has 3 components
  - Proposal (pass/fail)
  - Presentation (inadequate/poor/good/excellent)
  - Report (1-7)

For the visualization project, show that you can analysis, understand, and/or communicate or teach about data

- Multiple independent variables
- Multiple dependent variables
- Complex behavior over space
- Complex behavior over time

Chapter 1

Lecture Notes

1.1 Data Visualization

The use of images to provide insight into phenomena. Should reveal data:

- show the data, honestly
- thought-provoking (not distracting)
- efficient (many data in little space)
- encourage comparison
- expose comparison
- serve a purpose
- link closely to descriptive statistics/text

1.1.1 Visualisation Procedure

Iteractive process:

- Locate/acquire data
- Parse data
- Filter data
- Clean/analyse/derive
- Map to geometry
- Render
- Interact

1.1.2 Data acquisition

Access considerations:

- Need a reliable (credible) source (e.g. govt/university)
- Need the right to use the data
- Acknowledge source
- May need to register/pay
- May have to apply in writing
- Download directly/automatically?
- Dataset[s] may be huge/dynamic
- Can their server cope?
- Be a good internet citizen (... or get blocked)

1.2 Univariate data

Univariate data: multiple measurements for one thing

Bivariate data: multiple measurements of two things, *temperature and windspeed at a station*

Multivariate data: multiple measurements of 3 or more things

1.2.1 Descriptive Statistics

Measures of variation

Ranges: max-min, inter-quartile, boxplots

Standard Deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

Variance: s^2

Skewness: asymmetry $\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{\frac{3}{2}}}$ also (mean - mode)/s

Kurtosis: flatness (platykurtic) or sharpness (leptokurtic)

Types of errors in data:

- human and machine
- recording errors
- transcription/storage errors
- precision and rounding errors
- unit errors
- false presences/absences
- ... and so on

Two kinds of errors affecting all of our data:

Random error: This affects the **precision** of the data

Systematic error: This affects the **accuracy** of the data

1.3 Bivariate data

- **Paired** measurements of two quantitative variables/obervations
- could be just two variables, interested in their **relationship**
- or could be a response (y) to some factor (x)
- can still use univariate methods (quartiles, mean-differences, etc)

1.4 Time-series

Nature of Time series data

- unidirectional
- discrete/continuous/(oridinal?)

- point-based/intervals
- can be nested
  measure something every day, another dataset of the same measurement is taken hourly
- can exhibit **cycles**
  days, week(end)s, months, seasons
- some ideas may apply to other data with spacing, frequency

Time-series data can either discrete or continuous:

Continuous: temperature vs time

Discrete: rainfall per day

1.4.1 Time series periodicity

Fourier's theorem: Any periodic function of time can be expressed as a sum of sine and cosine functions (i.e. as a Fourier series). Not periodic? Then you get a continuous Fourier integral rather than a discrete Fourier series.

Fourier transform: Converts time-domain function to frequency-domain spectrum (Fourier series or integral, which we also call the Fourier transform).

Inverse Fourier transform: Frequency-domain back to time-domain.

Method used on the computer is known as a **Fast Fourier Transform (FFT)**.

1.5 Colour, light, and animation

1.5.1 Colour

- observation and interpretation of elements and relationships
- history and recommendations from cartography
- colour can:
  label
  measure
  represent reality
  emphasise
  enliven/decorate
- widespread
  but not trivial to get right

Rules

- good compromise: two hues, varying lightness
- keep strong colours for extremes
- not too many colours - 10 (paper), 15 (screen), 25 (greyscale)
- light/bright not next to white
- change hue with category,
- change saturation with rank/quantity
- avoid red/green contrasts

1.5.2 Animation

- attract attention, focus
- enjoyable, insightful
- enhance understanding
- great for complex objects
- worth the investment?
  time, effort, clarity (of graphics and info)

What can be bad about animation?

- It doesn't translate well to print
- It takes time and effort
- It can tie us to specific software
- It can make comparison harder - can you compare the current frame with a similar frame from 15 seconds ago?

Animation considerations

- Record/playback
  - large/complex surfaces
  - small set of stills, easily connected
- Real time animation
  - simple graphics objects
  - user interaction
- Other constraints
  - computer speed/memory
  - number of frames storable
  - complexity of animation
  - need for clarity not distraction (as always)

1.6 3D and 4+D Visualisation

- If we have 3 variables, we can plot them using 3 axes
- Gives us a 3D plot which is represented as a 2D image
- We may need cues to help identify the information
- More than 3D can be quite difficult to understand
  Quantitative methods become useful

1.7 Multivariate data analysis

Yeah look, there was content but idk

1.8 Spatial statistics

- **Location:** What's happening at positions of interest?
- **Pattern:** Spatial arrangement of phenomena/events
- Analogous to previous descriptive stats, + space

1.8.1 Measures of dispersion

Spectrum of dispersion (clustered $\rightarrow$ random $\rightarrow$ dispersed). Standard distance:

$$\sqrt{\frac{\sum_i d_i^2}{n}}$$

1.8.2 Nearest neighbour analysis

- Observed nearest-neighbour distance: D_{obs} = mean distance to all points' nearest neighbours
- Expected nearest-neighbour distance: $D_{\text{rand}} = \frac{1}{2\sqrt{p}}$
  $p =$ density of points
- max clustering: $D_{\text{obs}} = 0$
- max dispersion: $D_{\text{obs}} = \sqrt{\frac{2}{p\sqrt{3}}}$
- nearest-neighbour index $= \frac{D_{\text{obs}}}{D_{\text{rand}}} (\in [0, 2.15])$

1.9 Multivariate data analysis

Lets us reduce the number of dimensions (variables) we need to describe the data

1.9.1 Principal Component Analysis (PCA)

> PCA aims to let us describe most of the variation in our data with a small number of independent variables

1.9.2 Cluster Analysis

> Cluster analysis aims to classify the data into discrete groups. Then, we can describe a data point simply by specifying which group it belongs to

1.10 Visualising multidimensional data

1.10.1 Ternary plots

- Triangular ordination of *proportions* of 3 components (i.e. $\sum = 1$)
- tricky but worth the effort
- axes read parallel to each zero (center = 33% each)

1.10.2 Scatterplot matrix

- tiled 2d visualisation of multiple variables

1.10.3 Coplots

- conditional on one variable
  slice through data
  slices trade-off range and data
  fit curves to see dependency
  can also graph residuals etc.

1.10.4 3d point plots

- scatterplot in 3d
- used to seeing a 3d object in 2d
- if not too many points can use a stem plot

Interpolating

- if irregular x/y data
- create lattice covering range of x and y
- interpolate between z-values

viewing/presentation choices

- grid density
  - sparse = lose surface definition
  - dense = lose depth perception
- axis ratios
  - as data if axes share same units
  - otherwise 1-1-1
  - or banked z-axis?
- projection
  - perspective: good for real objects (but distortion)
  - orthogonal: fixed lengths $\rightarrow$ data analysis (but front/back confusion)
- orientation (azimuth and elevation)
  - multiple rotations helpful (around z-axis, in equal steps)
- other effects

  - box
  - ticks
  - curtain
  - show/hide rear wires
  - colour

All involve trade-offs
- help or hindrance
- should highlight data, not effect itself

colour representation

use of colour
- none
- height-related
- gradient-related

1.11 Computer Graphics (CG)

1.11.1 What is CG

- Algorithms, Techniques, and Processes that
- Execute on computers (by definition)
- Produce images (of some description)
  On screens, in files, VR headsets, printers, t-shirts, or anything really
- Related (sometimes very closely) but *not* CG as such
  - Image processing
  - Data acquisition / scanning
  - Fabrication (e.g. 3d printing)
  - Simulation
  - Visualisation

1.11.2 Offline

Higher quality image or series of images. Can take hours to compute which is fine because there is no human interaction.
- Frame time > 2 min
- Final image quality (render farms)
- Image stability (animation)
- Time to predictive image

1.11.3 Interactive

- 33ms - 1s/frame
- Performance (Special HW ok)

1.11.4 Real Time

Main goal is to be interactive, so quality is a secondary concern.

- <33ms / frame
- Single workstation performance

1.11.5 VR / AR

- <10ms / frame
- Single workstation performance
- Stereo images
- Wide field of view
- Low latency
- Disparity
- Image stability
- Close range detail

1.12 Coordinate Systems

1.12.1 Euclidean Space

Can be 2D or 3D and is defined as the "normal space", i.e. it behaves like the observable world. The space is made up of coordinates which are typically Cartesian

1.12.2 Points and Vectors

Point is a location or position, Vector has a direction and length (from two points)

1.13 Transformations

- Position/Movement: **Translation**
- Orientation: **Rotation**
- Size: **Scale**
- Inverse

1.13.1 Translation

- Movement along a vector
- For all points in space:
  $P' = P + t$
  $T(t)$ denotes translation by vector t

1.13.2 Rotation

- Rotation in angles around a vector
- Rotation around Y axis: $R_y(\alpha)$
- Rotation around arbitrary vector: $R(v, \alpha)$

1.13.3 Scale

Uniform, $S(s)$:

- Same in all axes
- For all points: $P' = sP$

Non-Uniform, $S(v)$:

- Different value for x, y, z
- For all points: $P' = vP$

1.13.4 Handedness

- Affects rotations
- Affects back-face culling
- A model is authored for LH or RH
  Mirrored if viewed in opposite
- Conversion *can* be tricky

1.14 Spaces