

Visualisierung

Aufgabe 1

Haris Becic, 0648038

Beschreibung der Aufgabe: Unsere Aufgabe beim diesem Uebungsbeispiel war zwei Visualisierungen (eine gute und schlechte) zu finden in diese in Dateils zu beschreiben. Wir sollten beschreiben was diese Visualisierung darstellt und uns auf die Sachen wie Zugänglichkeit, Klarheit, Genauigkeit usw. konzetrieren. Diese Visualisierungen sind in Textbooks, Journal, Science magazine newspaper zu finden.

Beschreibung der guten Visualisation:

Ich habe meine gute Visualisierung aus einem Artikel aus New England Journal of Medicine ausgewaehlt. Obwohl es einige interessante Visualisierungen aus dem Artikel (einschließlich einer guten Animation) sind, wählte ich die folgende Grafik(Figure 1) :

Link: <http://www.nejm.org/doi/full/10.1056/NEJMsa066082>

Was zeigt das Bild?

Diese Grafik zeigt ein soziales Netzwerk aus rund 2200 Personen aus Framingham Heart Study. Der Zweck von dieser Visualisierung war festzustellen ob es eine Verbindung zwischen Übergewicht und Person –to-Person Links gibt. In dieser Visualisierung stellt jeder Knoten eine Person dar. Die Farbe der einzelnen Knoten Grenze darstellt Geschlecht (rot: weiblich; blau: männlich), und die Größe des Knotens ist proportional zur der Person Body Mass Index (BMI). Die Autoren versuchen weiter zu kategorisieren ob eine Person entweder 1) fettleibig, oder 2) nicht fettleibig, je nachdem, ob ihr BMI oberhalb oder unterhalb 30. Diese Bezeichnung in der Grafik wird durch die Farbe der einzelnen Knoten codiert: gelb stellt eine übergewichtige Person, für grün eine Person, die nicht fettleibig. Die Farbe der Kanten bedeutet, die Art der Verbindung zwischen den Knoten: orange stellt eine familiäre Bindung und lila stellt Freundschaft oder Ehe. Die Position jedes Knotens wurde durch einen Algorithmus, um den Abstand zwischen jedem Knotenpaar zu beheben versucht erzeugt. Dies bedeutet, dass Knoten mit vielen Verbindungen eher in der Mitte des Netzes sind.

Kriterium Beschreibung:

Das Bild Modell ist eine Visualisierung des Netzwerks in einer 2D-Plot der Knoten (Düse Menschen) und sterben Kanten Zwischen den Knoten (Düse verbindungen Zwischen den Menschen). Die Größe der Knoten kodiert der Person BMI, sterben Farbe Westerwaldkreis.png es entweder in der fettleibig oder nicht-adipösen Kategorie, sterben Farbe der Kontur der Knoten Geschlechts codieren, wobei sterben Rander zwischen den Knoten zu einer Beziehung zwischen den Knoten (Personen) entsprechen, und die Farbe der Kanten zu einem entweder familiärer oder Freundschaft Verbindung korrespondieren. Knoten Position (x-Achse und y-Achse) nicht direkt codieren irgendwelche Daten, aber es indirekt kodiert, wie viele Verbindungen ein Knoten (Knoten mit vielen Verbindungen Neigen dazu, in der Mitte seins can). Grundsätzlich ist die Anordnung von Knoten automatisiert unter Verwendung eines Algorithmus, um sterben Komplexität des Netzes Visualisierung zu reduzieren, und sterben Positionen sind einfach ein Nebenprodukt von diesem Prozesses.

Warum?

Ich wählte diese auf meine gute Visualisierung , weil es zeigt den Nutzen der effektive grafische Darstellungen für große und komplexe Daten zu denken. In dieser Studie, wäre es fast unmöglich, eine allgemeine und unmittelbare Sinn der Daten, insbesondere die Verbindungen zwischen Teilnehmern und dem Muster im Zusammenhang mit Fettleibigkeit, ohne Ausnutzung der perzeptuellen Bandbreite des menschlichen visuellen Systems zu erhalten. Die schiere Größe des Datensatzes würde es viel schwieriger und zeitaufwendig, um die Daten inhaltlich zu untersuchen oder untersuchen die Zusammenhänge zwischen verschiedenen Variablen in einer stückweisen Weise.

Diese Art der Visualisierung ist besonders nützlich, als ein Werkzeug für die Erkundung der Daten und suchen nach Mustern, die entstehen, wie verschiedene Attribute optisch kodiert werden kann. Ohne diese Grafik, könnte der Suche nach einem Muster in den Daten möglicherweise nehmen mehrere Stunden mit unzähligen Analysen. Angesichts dieser Visualisierung ist jedoch der Zusammenhang zwischen Übergewicht und sozialer Verbindungen relativ leicht zu sehen, wie man die verschiedenen Gruppen von ähnlich aussehenden Knoten untersucht. Diese Art der Untersuchung der Daten kann einem helfen, genau feststellen, welche weiteren quantitativen Analysen am ehesten zu signifikanten Ergebnissen führen können. In diesem 2D-Grafik macht die Kompaktheit der Knotenpositionen es einfacher, die gesamte Teilkomponente zu sehen, aber dies kann zu einem Problem mit einem Verschluss führen. Es ist nicht immer klar, wie verschiedene Knoten im Zentrum des Netzes in dem viele Knoten und Kanten überlappen miteinander verbunden sind.

Beschreibung der schlechten Visualisation: Ich entschied mich für mein schlechtes Visualisierung aus einer Lehrbuch der Psychologie, die in einem Bachelor-Intro to Psych natürlich verwendet wurde:

Link: <http://www.davidmyers.org/Brix?pageID=102>

Was zeigt das Bild?

Die obige Grafik wurde in einem Abschnitt des Buches widmet sich der Forschung schlafen vorgestellt. Das Balkendiagramm fasst Daten, die die Zahl der Verkehrsunfälle in Kanada vor und nach Sommer- und Winterzeit Anpassungen für die Jahre 1991 und 1992 (kombiniert). Der Zweck dieser Grafik ist, eine Korrelation zwischen verlorenen Schlaf und Verkehrsunfällen vorzuschlagen. Ganz allgemein denke ich, die Absicht ist, einfach zeigen, dass ein Mangel an Schlaf kann die körperlichen und geistigen Fähigkeiten mit sicherem Fahren verbundenen beeinträchtigen.

Kriterium Beschreibung:

Die Grafik zeigt die Unfallhäufigkeit nach unterschiedlichen Zeitraum Kategorien. Diese Visualisierung basiert auf einem statistischen Datenmodell, in dem Unfallhäufigkeit höchstwahrscheinlich durch ganze Zahlen, die die Anzahl von Unfällen an einem bestimmten Tag des Jahres codieren vertreten ist. Aus diesem Diagramm die Autoren nur untersucht bestimmten Zeitraum Kategorien (nominal Etiketten). Die Daten bestehen aus zwei Hauptkategorien unterteilen: die Feder Zeitemstellung (eine Stunde zu verlieren) und die Abfallzeit Änderung (gewinnt eine Stunde). Jede dieser Kategorien wird weiter in Pre- und Post-Zeitumschaltung Kategorien unterteilt. Auf einer bestimmten Ebene wirken diese Kategorien einfach nominal Etiketten, aber es ist ein wichtiger Bestellung bis ihnen (vorher und nachher) sowie eine Richtung (gewinnt vs eine Stunde zu verlieren), und die Daten können nur dann richtig mit diesen Überlegungen werden im interpretierten Geist.

Das Bild basiert auf univariate Daten basieren, und die Autoren kodieren die Unfallhäufigkeit Preise als einfaches Balkendiagramm. In dem Graphen liegen die Kategoriebeschriftungen entlang der X-Achse und der Unfall Frequenz entlang der Y-Achse aufgetragen. Die Grafik untersucht nur Preise für die Montags vor und nach der Zeitemstellung, wahrscheinlich, um sicherzustellen, dass der Tag der Woche ist nicht ein Faktor in jedem gemessenen Effekt. Da es sich um ein Balkendiagramm, die Höhe der Balken auf der Unfallrate entsprechen. Darüber hinaus werden die Stäbe nach den Pre- und Post-Zeitumschaltung Kategorien farbkodiert, vermutlich, um es einfacher, die verschiedenen Bedingungen zwischen Frühjahr und Herbst zu vergleichen.

s gibt mehrere Gründe, warum ich glaube, diese Grafik nicht ein genaues Bild von den Daten geben. Zuerst werden die zwei Hauptkategorien (Frühjahr, Herbst) auf separaten Diagramme, von denen jeder einen anderen Bereich auf der Y-Achse hat dargestellt. Die max für den Frühling Veränderung ist 2800, während der Rückgang Zeitumstellung maximale Wert ist 4200 (obwohl dies oben Teilstrichbeschriftungen tatsächlich unterhalb der oberen Teilstrichbeschriftungen für das Frühjahr positioniert). Es gibt mehr Unfälle während der Herbst-Zeitumstellung insgesamt noch sowohl die Pre- und Post-Daten Bars aus dieser Kategorie sind optisch kleiner als die jeweiligen Bars im Frühjahr Kategorie. Ebenso kreuzen die y-Achse die x-Achse bei verschiedenen anderen Werten als Null: die Feder chart beginnt bei 2400 Unfällen, während der Rückgang chart beginnt bei 3600. Mit nur Visualisierung einen kleinen Teil der Daten kann es zu Unterschieden erscheinen auffällig (und größere) trotz der Tatsache, dass sie nur für eine kleine prozentuale Veränderung der Gesamtzahl der Unfälle. Dies macht es auch schwierig, die beiden Diagramme direkt vergleichen.

Noch wichtiger ist, stellen die Tick-Mark für die beiden unterschiedlichen Diagrammen verschiedene Bereiche. Obwohl die beiden Diagramme Seite-an-Seite angezeigt werden, um Vergleiche zu ermöglichen, jeweils Teilstrichbeschriftungen auf der linken Diagramm 100 Unfälle während eine ähnliche Teilstrichbeschriftungen rechts entspricht 200 Unfälle. Der physikalische Abstand der Markierungen für gleiche Abstände auf Papier, doch entsprechen unterschiedlichen numerischen Mengen in den Daten.

Ich denke, ein weiteres Scheitern des Graphen, die in irgendeiner Weise auf einen Mangel der Analyse im Allgemeinen verwandt ist, ist, dass es keinen Kontext, in dem diese Daten zu sehen. Wir sehen nur Unfallraten am Montag vor dem Zeitpunkt Veränderung gegenüber dem Montag nach der Zeitumstellung, und nur für einen kombinierten Zwei-Jahres-Zeitraum. Obwohl die Autoren ausdrücklich, wie Unfallraten als Folge der Sommerzeit ändern interessiert sind, wäre es immer noch nützlich sein zu sehen, wie Unfallraten auf wöchentlicher Basis durch ein ganzes Jahr zu ändern. Ebenso kann es sinnvoll sein, zu sehen, wie diese Sätze im Laufe von mehreren Jahren zu ändern. Die Angaben in dieser Grafik erscheinen die Autoren das Argument unterstützt, aber diese Arten von Änderungen können häufig auftreten im Laufe des Jahres und das Timing kann Zufall sein. Wir können auch nicht sagen, ob dieser Trend von Jahr zu Jahr wiederholt wird. Die Macht der Kurve wird durch diesen Mangel an Kontextinformationen geschwächt.

Schließlich umfasst die Visualisierung der Balkenanzeige einige unnötige (und sogar ein wenig verwirrend) 3D Styling, trotz der Tatsache, dass es nur zwei Dimensionen auf die Daten (Unfallrate und Zeiträume). Konkret wirft jeder Balken in der Grafik einen leichten Schatten auf den Hintergrund. Mit dieser zusätzlichen Dimension, ist es schwierig, die genaue Höhe der Balken messen (wir verwenden die Bars oder die Schatten, oder irgendwo dazwischen). Höchstwahrscheinlich wurde die Schattierung für Stil allein aufgenommen, aber es dient nur dazu, ein wenig mehr Unklarheit zu einem bereits schwachen grafischen Darstellung hinzuzufügen.