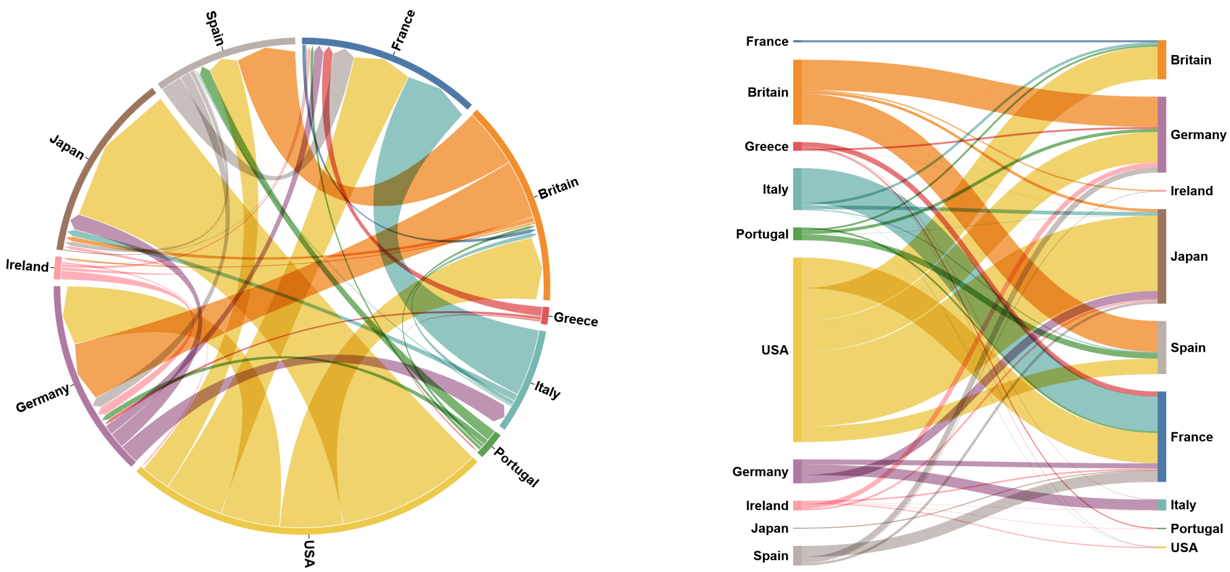
Check for updates

Showing Flow: Comparing Usability of Chord and Sankey Diagrams

Carl Gutwin
Department of Computer Science
University of Saskatchewan
Saskatoon, Canada
gutwin@cs.usask.ca

Aristides Mairena
Department of Computer Science
University of Saskatchewan
Saskatoon, Canada
aristides.mairena@usask.ca

Venkat Bandi
Department of Computer Science
University of Saskatchewan
Saskatoon, Canada
venkat.bandi@usask.ca



(a) Chord: arrows lead from debtor to lender.

(b) Sankey: debtors shown are at left.

Figure 1: Chord and Sankey diagrams showing the debt dataset.

ABSTRACT

Chord and Sankey diagrams are two common techniques for visualizing flows. Chord diagrams use a radial layout with a single circular axis, and Sankey diagrams use a left-to-right layout with two vertical axes. Previous work suggests both strengths and weaknesses of the radial approach, but little is known about the usability and interpretability of these two layout styles for showing flow. We carried out a study where participants answered questions using equivalent Chord and Sankey diagrams. We measured completion time, errors, perceived effort, and preference. Our results show that participants took substantially longer to answer questions with Chord diagrams and made more errors; participants also rated Chord as requiring more effort, and strongly preferred Sankey diagrams. Our study identifies and explains limitations of the popular Chord layout, provides new understanding about radial vs. linear layouts that can help guide visualization designers, and identifies possible design improvements for both visualization types.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**;

KEYWORDS

Chord diagrams, Sankey diagrams, radial vs. linear layout, visual analytics, visualization usability

ACM Reference Format:
Carl Gutwin, Aristides Mairena, and Venkat Bandi. 2023. Showing Flow: Comparing Usability of Chord and Sankey Diagrams. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3544548.3581119

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI '23, April 23–28, 2023, Hamburg, Germany
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581119

1 INTRODUCTION

Many visualizations show connections and flow, and designers have several choices when considering how to represent this kind of data. Two common techniques for showing flow are Chord diagrams, which use a radial layout, and Sankey diagrams, which use a left-to-right layout with two axes. Although both types are commonly used, Chord diagrams have recently become popular both in mainstream media (e.g., [46]) and in research communication (e.g., [35, 39, 48]). Chord diagrams can be produced with several visualization tools such as D3 [5], R [24], PowerBI [17], Bokeh [29], and ArcGIS [42]. Despite several previous studies of radial visualizations [1, 9, 11, 14, 16, 32, 49], there is still little understanding of how radial and linear presentations compare when used with flow data. The design differences between Chord and Sankey diagrams can substantially change the way that users extract information from the visualization: e.g., Chord's radial layout requires that people mentally rotate entities, does not provide an obvious overall direction for flows, and makes heavy use of the circle's centre for links. To test interpretation performance with Chord and Sankey visualizations, and to add to the comparison of radial and linear layout approaches, we carried out a crowdsourced study where participants (N=51) used equivalent Chord and Sankey diagrams for five tasks (determining the existence of a connection, finding entities that satisfy criteria, comparing magnitudes, finding extreme values, and counting links) in four different datasets. We measured the time needed to answer questions and the number of errors, as well as participants' subjective effort ratings and overall preferences. Overall, our results showed performance and usability advantages for the left-to-right Sankey diagrams:

- Completion-time results showed that participants took longer to answer all question types with Chord diagrams (3.7 seconds per question); the difference in completion time was also much higher for the first iteration of each question type (Chord was more than 9 seconds slower on average for the first iteration).
- Accuracy results showed that participants also made more errors with Chord than with Sankey (0.1 more errors per question on average); again, there was a larger difference on the first question iteration (0.42 more errors).
- Participants reported higher effort with Chord diagrams on the NASA TLX scale, with Chord seen as requiring higher mental effort and overall work, causing higher frustration, and having reduced performance.
- Sankey diagrams were strongly preferred by participants (42 of 51 people chose Sankey overall).

The differences between the two techniques are large enough to have a real-world effect on usability, particularly when people carry out a new type of query. The congruence of several measures (time, errors, effort, and preference) suggests that there are real differences in how users perceive and interpret the two techniques. We discuss our results in terms of the fundamental design and layout differences between the two techniques.

Our work makes four contributions. First, we provide empirical evidence about the usability and interpretability of two common visualization techniques for showing flow, and we identify advantages for linear diagrams in terms of time, errors, effort, and preference. Second, we summarize design factors that must be considered when developing either Chord or Sankey diagrams. Third, we add to the ongoing assessment of radial versus non-radial approaches, and show additional limitations of radial layout. Fourth, we provide possible design improvements for both visualization types.

2 RELATED WORK

2.1 Visualizations of Connection and Flow

Many styles of visual representation have been proposed to show flows between entities. These can be organized into two main groups: representations that focus primarily on the entities, links, and directions of flow; and those that also consider flow magnitudes or amounts. The first group contains well-known representations such as node-link diagrams (e.g., a software call graph [15, 27]) or flow charts (e.g., a workflow for diagnosing hardware errors [23]).

The second group represents flow *amounts* in addition to entities, links, and directions – often encoded using the width of the link between related entities. Sankey diagrams and Chord diagrams fit into this group. Chord diagrams lay out entities around a circular axis, and represent weighted connections or flows by linking entities across the circle. Chord diagrams have been used in many application domains including bioinformatics [35, 36], engineering [25], medicine [41], anthropology [3], and management [26]. Sankey diagrams have also been used in many different domains to show the flow of energy, money, material, or people (see numerous examples at www.sankey-diagrams.com). Both diagram types represent entities, links, directions, and flow amounts, with the main difference in their layout (Chord using radial and Sankey using left-to-right linear). Various studies have compared these two general layout strategies across chart types, and strengths and weaknesses for comprehension have been identified, as discussed below.

2.2 Radial and Non-Radial Organizations of Visual Data

Many types of visualizations can be laid out either using a radial or a linear approach [10], and several studies have been carried out to compare performance with the two layouts. These studies have shown both benefits and drawbacks for radial layout (with contrasting evidence about some features) [1, 9, 11, 14, 16]. In the 1980s, Cleveland and McGill found that pie charts are less effective than equivalent bar charts due to difficulty in comparing angles [11]. Later eye-tracking studies also showed reduced performance for radial charts compared to non-radial layouts: e.g., in scatterplot designs, participants were slower when mapping data points to values with radial axes [19]. Even when the radial layout matches a real-world analogue (such as a clock), interpreting radial diagrams can be difficult: studies have found that interpreting patterns was slower with a radial layout that matched a clock face, compared to bar charts [49] or flower and circle charts [1], with variants such as sunburst charts being least preferred [54]. Other studies have shown reduced memorability of locations in radial presentations (e.g., participants made more errors in remembering radial sectors than in remembering grid locations [14]) and increased time to read spiral and circular timelines compared with linear versions [13].

Several benefits of radial designs have also been proposed and evaluated in prior work. Radial layouts can reduce the length of connections (e.g., a wrap-around layout allows links to go in either direction), potentially reducing clutter (e.g., [8, 10]). Similarly, circular concentric layouts provide different amounts of space per ring, which can be useful when some data (e.g., the most recent time period in a time series chart) needs to be shown in more detail [10] or on smaller screens [6]. If the dataset has a periodic nature, radial presentations can provide a natural method for comparing changes over time (e.g., yearly rings in a spiral chart [51]). Finally, several authors have noted the aesthetic appeal of radial diagrams (e.g., [10, 14, 32]), although evidence is limited to participant comments indicating a preference for radial designs (e.g., [49]).

2.3 Comprehension and Perception of Visualizations

Visual characteristics can influence how attention is guided to specific points through the use of color, shading, and animation, primarily from factors arising from graphical perception (see [21] for a review). Three perceptual factors are of particular importance given the design differences between Chord and Sankey diagrams: interpretation of rotated items on circular axes; visual search in cluttered spaces; and visual path following.

Rotation. Several researchers have considered the costs of interpreting text and objects at non-horizontal orientations (e.g., [31, 47]). This work suggests that people must mentally rotate visual elements in order to interpret or compare them, and that there is a direct relationship between the change in orientation and the cost of interpretation [52].

Visual search. Line-based clutter is of particular importance for flow diagrams, and studies show that crossings can pose problems for interpretation (e.g., [30]). Early perceptual filters (i.e., edge detectors) are relatively coarse, meaning that acute-angle crossings are more confusing than crossings that are close to 90 degrees [50].

Path following. Following connections in a flow diagram involves visual tracing of the link between entities. The speed of visual tracing has been shown to be dependent on the path length and curvature as well as the proximity and crossings of distractor paths. Longer lines require more tracing time [28], path segments that are on a smooth curve are easier to see, and the degree of "bendiness" correlates with the time required to interpret length [50].

3 DESIGN FACTORS FOR CHORD AND SANKEY FLOW DIAGRAMS

Chord and Sankey diagrams show weighted network datasets with entities represented as nodes and connections as links. Many of the characteristics of the diagrams are determined by their radial or left-to-right setup, but there are several decisions that designers still need to make regarding layout, encoding, routing, and labels.

- *Layout and encoding of entities.* Entities must be laid out on the single or multiple axes of the flow diagram, and links from each entity are often encoded using colour. Layout of entities may follow the inherent ordering of the dataset (e.g., chromosome number in genomes); if this is not a constraint then the entities can be rearranged to reduce link crossings [43]. Entities in Chord diagrams can be source nodes, destinations, or both; in a Sankey diagram, source entities are on the left and destination entities on the right (entities can appear on both axes).
- *Layout of links.* Links may have weights, often encoded using width (although other visual variables could be used); designers may impose a minimum width in order to ensure visibility. Sankey and Chord diagrams provide different space for attaching links to entities: Chord uses the circle's circumference, and Sankey uses the two axes. This means that there is more space available in a Chord diagram (about 1.4x for diagrams with the same area); taller Sankey diagrams increase link space (at the cost of link-routing space).
- *Link crossings.* Visual clutter from crossing links can be a problem for both diagram types. Designers can paint smaller links on top of larger ones to ensure visibility, or can use transparency to show underlying links. If nodes do not have a fixed order, crossings may be reduced by changing the layout of the nodes.
- *Link direction.* Link direction is often encoded using color (of the source node) and with an arrowhead at the destination end (e.g., D3's *ribbonArrow* component). Arrowheads are typically incorporated into the ribbon to avoid changing the maximum width of the ribbon. If an entity is both a source and a destination in a Chord diagram, incoming and outgoing links are often grouped; the repeated arrows provide a visual distinction between incoming and outgoing links. In addition, links are often sorted by link width. In Sankey diagrams, link direction is implied by the left-to-right layout, so arrowheads are not required.
- *Label orientation.* In a Chord diagram, labels can be oriented radially, horizontally, or circularly. Previous research has shown that rotated text is more difficult to read [52], but horizontal labels may not be easy to achieve for Chord diagrams if several entities are clustered at the top or bottom of the circle. Many libraries including D3, Circos, and Bokeh use radial layout by default. For Sankey diagrams, labels can always be horizontal.
- *Aspect ratio.* Sankey diagrams can use different aspect ratios: if there are many nodes, the diagram can be stretched vertically; if there are complex links, the diagram can be stretched horizontally to better lay out links.

4 STUDY METHODS

4.1 Design of Chord and Sankey Visualizations

We developed a custom Javascript library to enable our comparisons of equivalent Chord and Sankey diagrams. The library uses D3.js [5] for link routing and ribbon specification, and renders diagrams to SVG for presentation in a web page. The diagrams were designed using the principles described above and following practices seen in previous work (e.g., [32, 35, 39, 46]). Chord diagrams used a single chord for each entity and grouped incoming and outgoing links at that node, with links ordered by width. Link routing and shape was controlled by the D3 chord package, using the default control points, and arrowhead parameters as specified by D3's 'ribbonArrow' command. We used D3's default orientation for labels (drawn radially outward from the entity). Sankey diagrams placed entities on the source and destination axes, with link routing using default D3 control points. In both diagrams, nodes were given a unique color (using D3's *Tableau10* palette) that was also used for outgoing links. In the Space dataset, links are unweighted, so ribbons in both diagrams were given a fixed width. Diagrams had equal areas (not counting text labels): on a 27-inch 1080p monitor, Chord diagrams were 11cm circles, and Sankey diagrams were 8cm x 11.5cm rectangles. Both diagrams ordered the entities based on the organization of the source dataset; we did not rearrange entities for any diagram. We inspected all diagrams to ensure that neither diagram type had noticeably more link crossings than the other (see Figures 1a-7). We did not add any interactivity (e.g., hover highlighting) to either type, as a main goal of the study was to assess the diagrams' ability to support purely visual interpretation and information-seeking.

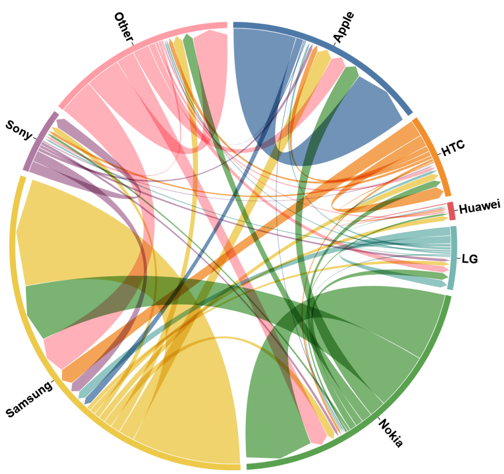


Figure 2: Chord: phone dataset. Arrows show brand switch.

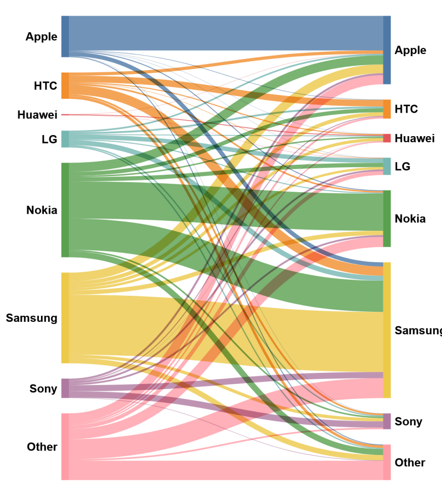


Figure 3: Sankey: phone data. Previous brand at left.

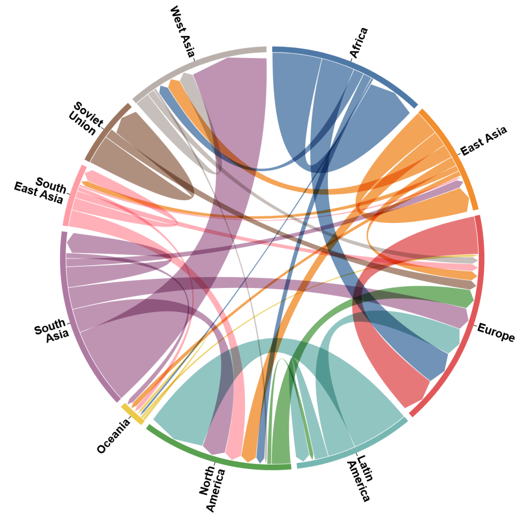


Figure 4: Chord: immigration dataset.

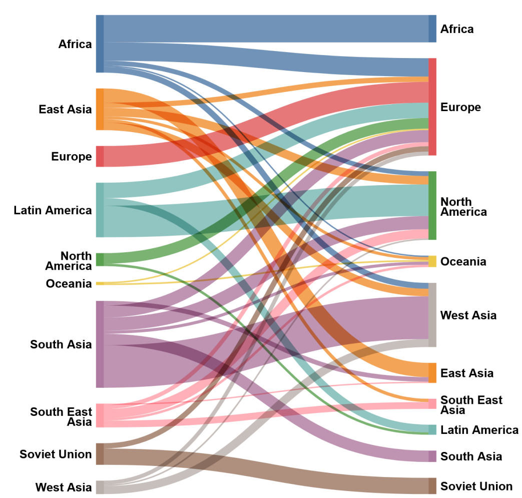


Figure 5: Sankey: immigration ('from' country at left)

4.2 Tasks and Datasets

We chose four datasets and five types of questions for the study that could be used with either Chord or Sankey diagrams. Our question types draw from Amar's set of low-level components for analytic activity [2] and the extensions proposed by Quadri and Rosen [40]. We chose questions that involved realistic assessments that would be carried out when doing analytic work with flow diagrams, and that could be answered through inspection of the visual features of the diagram (entities, connections, flow directions, and flow amounts) without needing extensive knowledge of the domain. In addition, we chose questions that would work equally well with both diagram types, and that could be answered in a reasonably short time. All questions were checked to make sure that they could be successfully answered with both Chord and Sankey representations (e.g., no questions involved lines smaller than eight

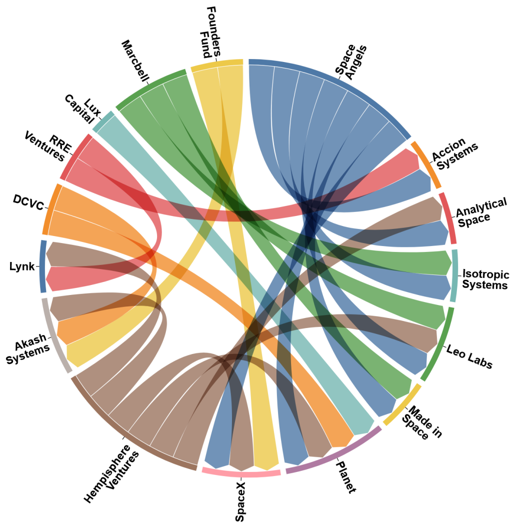


Figure 6: Chord: space investment dataset.

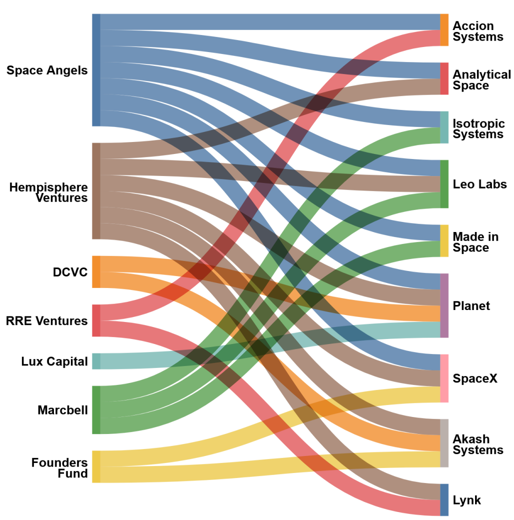


Figure 7: Sankey: space investment dataset.

pixels in width, links with similar colours, or highly-cluttered areas); in addition, questions were checked to ensure that they were fair to both representations (i.e., no questions involved obvious anomalies that would disadvantage either visualization). The question types are as follows (with examples from the immigration dataset as shown in Figures 4-5):

- *Existence*: determine if an entity or link exists in the diagram. Example: Is there a link from Africa → East Asia?
- *Find element*: find an element that matches certain criteria such as weight or number of connections (a version of Amar's "Filter" task [2]). Example: What region receives the most people from South Asia?
- *Comparison*: find two elements of the diagram and compare them in terms of certain criteria. Example: Which is larger: S.E. Asia → N. America, or Latin America → Europe?
- *Maximum / Minimum*: find the entity or link that has a largest or smallest value (similar to Amar's "Find Extremum" task [2]). Example: Which region has the largest link going to West Asia?
- *Count Links*: identify the number of links (incoming or outgoing) that are related to a particular node. Example: How many links go to South-East Asia?

We developed two question sets for each dataset since we used a within-participants study design (question sets were counterbalanced in the study). We also used four datasets in the study that cover a range of topics; all use publicly-available sources, and all have previously been used to demonstrate Chord diagrams (e.g., [4, 7, 20, 48]).

- *Debt*: shows debt imbalances during the 2010 "Euro debt crisis." Links connect debtor and lender countries, with link weights indicating the money owed (Figs. 1a and 1b).
- *Immigration*: global bilateral migration flow (1960-2015) between ten geographic regions; links connect countries, with weights indicating the number of immigrants (Figs 4 and 5).
- *Phone*: data from a survey in the Netherlands in 2000 about people's current and previous phones; links represent customer movement between brands (Figs 2 and 3).
- *Space*: connections between investment companies and companies in the space industry. This dataset has no link weights, so links are drawn with a fixed width (Figs 6 and 7).

4.3 Apparatus and Procedure

We developed a web-based study system to present diagrams, multiple-choice questions, and questionnaires. The study was administered through Amazon's Mechanical Turk infrastructure. Participants completed an informed consent form and a demographics questionnaire; they were then introduced to the two diagram types and completed a set of practice questions (with a different dataset) for each diagram (all question types were included). Participants were then assigned to an order condition (Chord or Sankey first), and a Latin-square group for datasets / question sets.

In the main study, participants were shown a visualization (Figure 1a – Figure 7) and a text description that clearly indicated how to interpret flow direction. Participants were then asked five questions (one of each type) with multiple-choice answers; participants selected a radio-button item from a list and pressed a Submit button to answer. If participants selected an incorrect answer, the system asked them to try again; once the correct answer was chosen, the system moved to the next question. After answering all five questions with one visualization type, participants then moved to the other visualization and answered five different questions (again, one of each type). Once questions were completed for the second visualization, the participant moved to the next dataset and repeated the process. After all four datasets, participants completed a NASA Task Load Index (TLX) survey for each visualization type, and answered questions regarding their perceived speed and accuracy with the two visualization types, as well as their overall preference.

4.4 Participants

57 participants were recruited through Amazon's Mechanical Turk, an online platform where workers can complete tasks for fixed remuneration. Several previous HCI and visualization studies have successfully used data gathered through MTurk [22, 33, 44, 45], with the caveat that causes of outliers such as bots and negligent workers must be accounted for in the analysis. To ensure the integrity of our data, participation in our study required workers to have over 90% task acceptance rate (a measure of worker quality used in previous work). We also checked survey responses to ensure that the same answer was not used for all questions, and whether questions were being answered too quickly. Ethical approval for the study was provided from the ethics board of University of Saskatchewan.

Six participants did not complete the study, leaving 51 complete sessions (32 men, 19 women, mean age 37.2). All participants were experienced using desktop computers (more than 20 hours/week) and 47 of 51 reported using some form of visualization in an average month (bar, pie, and line charts; scatterplots; data dashboards; schematic diagrams; and maps). None of the participants reported using any type of flow visualization (including Chord or Sankey diagrams). No participants had to be removed for answering the questionnaires too quickly or using the same response for every question. All participants were paid $6 USD; the study took approximately 30 minutes.

4.5 Study Design

Our goals were to compare radial Chord and linear Sankey visualizations in terms of time and errors in answering questions, people's perception of effort, and their overall preferences. We were also interested in whether any differences between the visualizations were consistent across different question types and datasets, and whether differences changed as users gained experience. This led us to the following within-participants factorial design:

- *Visualization Type (VisType)*: two levels (Chord and Sankey);
- *Question Type*: five levels (Existence, Find Element, Compare Magnitude, Min / Max, Count Links);
- *Dataset*: four levels (Debt, Immigration, Phone Switching, and Space Investment)
- *Question Iteration*: four levels (first, second, third, or fourth time the participant saw each question type); note that this factor is derived from the other combinations.

Our dependent measures were completion time and error counts per question, subjective ratings of effort (NASA TLX), and preference choices. The order of VisType and Dataset were counterbalanced (VisType was fully counterbalanced and Dataset used a Latin Square design). Participants worked with both visualizations within each dataset (e.g., if their visualization order was Chord → Sankey, and they were in the Debt dataset, they answered all five questions with Chord, then five questions with Sankey, before moving on to the next dataset). The questions were always presented in the same order (Existence, Find Element, Compare Magnitude, Minimum / Maximum, Count Links). Each participant therefore completed 2x5x4=40 questions; with 51 participants, a total of 2040 datapoints.

5 RESULTS

Analyses are organized below by dependent variable (completion time, errors, perceived effort, and preference). The effect size for significant ANOVA results are reported as generalized eta squared (η^2) [38], with < .01 considered small, .06 medium, and > .14 large [34]. Follow-up t-tests were corrected using the Holm-Bonferroni method. Outliers (trial time and errors) were capped based on pilot testing (i.e., maximum reasonable values needed to successfully answer the questions). Times were capped at 180 seconds, and error counts at 6 errors. Ten of the 2040 trials were capped for completion time, and 98 of the 2040 trials for error counts. Other than the six participants who did not complete the study, no participants were removed for other data-quality issues.

5.1 Completion Time

Our completion time measure recorded the time needed for participants to correctly answer the questions, including time to fix errors. Over all trials, mean completion time for Chord was 22.01 seconds, and for Sankey was 18.35 seconds (a difference of 3.74s). Figure 8 shows completion time by visualization type and question type, and Figure 9 shows these results separately for each dataset. We carried out a 2x4x5 RM-ANOVA to test for the effects of three factors (VisType, Dataset, and QuestionType) on completion time. The ANOVA showed significant main effects of VisType ($F_{1,50} = 16.93, p < .001, \eta^2 = 0.02$) and QuestionType ($F_{1,50} = 4.83, p < .05, \eta^2 = 0.0098$), but not Dataset ($F_{3,150} = 0.19, p = .90$). There were no interactions with VisType (all $p > .05$).

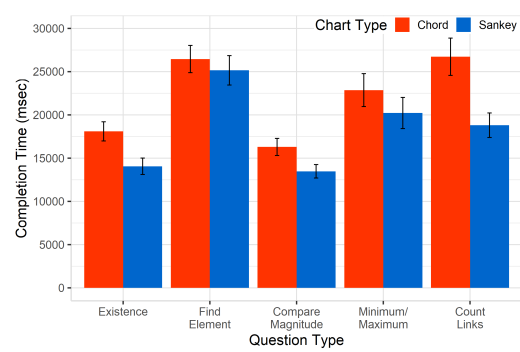


Figure 8: Completion time (±s.e.) by question type.

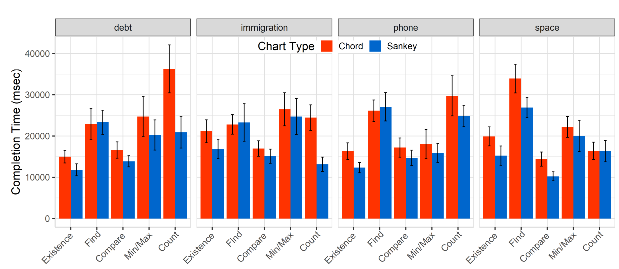


Figure 9: Completion time (±s.e.) by question type & dataset.

We also carried out a second analysis of completion time in terms of participants' increasing familiarity with the representations over the course of the study. For this analysis we used the question iteration as a factor – that is, whether it was the first, second, third, or fourth time a participant answered questions of each type. Participants saw four question iterations (which included one question of each type) for each visualization. (Note that the four datasets were seen in different orders for different participants, so each iteration includes all datasets).

We carried out a 4x2 RM-ANOVA to test for additional effects of QuestionIteration and VisType on completion time. The RM-ANOVA showed a significant main effect of QuestionIteration ($F_{3,150} = 7.47, p < .001, \eta^2 = 0.03$), as well as an interaction between QuestionIteration and VisType ($F_{3,150} = 4.89, p < .005, \eta^2 = 0.018$). We found participants improved as they gained more experience with the visualizations – but there were clear differences between the visualizations, with Chord substantially slower (compared to Sankey) on the first question iteration than on iterations 2-4. By the fourth iteration, there is little difference between the two types.

5.2 Errors

Errors were measured as the number of incorrect responses to each question. Over all trials, Chord had 1.07 errors per question and Sankey had 0.96. Because different questions had different numbers of choices, however, error counts differ substantially by question type. Figure 10 shows errors per question by visualization type and question type, and Figure 11 shows these results separately for each dataset. We carried out a 2x4x5 RM-ANOVA to test for the effects of VisType (in conjunction with Dataset and QuestionType) on errors. The ANOVA showed a significant main effect of VisType ($F_{1,50} = 5.37, p < .05, \eta^2 = 0.001$); there was no effect of Dataset ($F_{1,50} = 0.47, p = .71$), and we did not look for a main effect of QuestionType because of the differences in each type's number of choices. There was also an interaction between all three factors ($F_{12,600} = 2.23, p < .01, \eta^2 = 0.007$); as shown in Figure 11, the three-way interaction appears to be caused by particularly high error counts for Chord with "count links" questions.

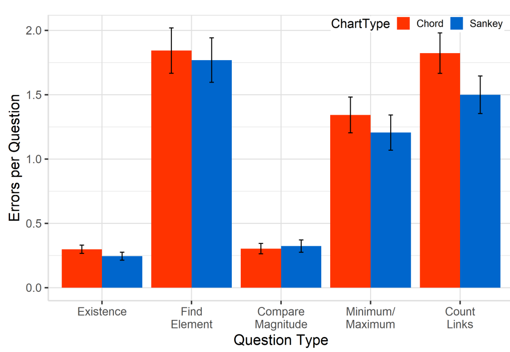


Figure 10: Errors (±s.e.) by question type.

As with completion time, we also considered the effect of increasing familiarity on errors, using question iteration as the main factor. We again carried out a 4x2 RM-ANOVA (QuestionIteration and VisType); no main effect of QuestionIteration was found

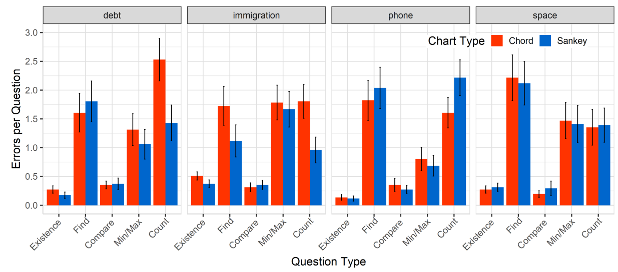


Figure 11: Errors (±s.e.) by question type and dataset.

($F_{3,150} = 0.92, p = .43$), but there was an interaction ($F_{3,150} = 5.24, p < .005, \eta^2 = 0.015$). Our results show substantially more errors with Chord visualizations in the first and second iterations (particularly in the Debt and Immigration datasets); but as participants became more experienced, errors improved for Chord.

5.3 Perceived Effort, Preferences, and Participant Comments

After each visualization type, people completed a NASA TLX questionnaire to assess their perceptions of effort, frustration, and performance. Mean scores are shown in Figure 12. After applying the Aligned Rank Transform [53], one-way ANOVAs were performed on each question, using VisType as the factor. We found significant differences between Chord and Sankey in terms of mental effort, perceived success, the amount of work required, and the frustration and annoyance caused (all $p < 0.01$). In all cases, the tests favoured Sankey over Chord.

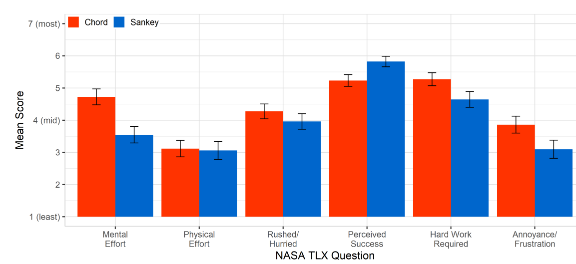


Figure 12: Perceived effort scores (±s.e.) (NASA-TLX).

At the end of the study, once participants had worked with both visualizations, we asked which they thought was the fastest, which was the most accurate, and which they preferred overall. For all questions, participants strongly preferred Sankey over Chord (Table 1): 35 of 51 felt that they were faster with Sankey, 43 of 51 felt that Sankey was more accurate, and 42 of 51 preferred Sankey overall. Chi-squared tests were all significant (all $p < 0.05$).

Preference Question	Chord	Sankey	χ^2	p
Which was fastest?	16	35	7.08	$< .05$
Which was most accurate?	8	43	24.02	$< .001$
Which did you prefer overall?	9	42	21.35	$< .001$

Table 1: Preference questions and results.

Participants' comments raised several issues regarding the designs. Several people stated that they found the Sankey layout to be easier to understand in general: one person said that Sankey "made it much easier to see certain things ... it was much less confusing"; another stated that Sankey "seems less complicated, easy to read"; another said "I found Sankey the easiest to read and understand. It was less mentally exhausting to look at"; and a fourth stated "It seems easier to look at for me." Participants also talked about how specific design features factored into their preference. Several people stated that they found it easier to follow links in a left-to-right layout: e.g., "It's just easier to read left to right", "the Sankey model was easier to read because of the unified flow (from left to right) which made the whole visualization less confusing, more organized", and "Sankey was a whole lot easier to follow as it only went left to right." Other participants focused on the link layout in Chord: one participant stated "The chord diagram seemed very all over the place. It was genuinely hard to follow as everything was going off in all directions"; another said "I found Sankey much easier to read rather then trying to detangle a pie chart of colors". Second, people commented on the problem of visually tracing links. One participant stated "It was easier to read and follow the lines [in Sankey]"; another said "The lines are a little less intersecting and easier to pin point when they get small"; and a third stated "It took me longer to be able to track data from one point to another on the circle [...] The colors helped, but it still wasn't as easy." Third, a few participants mentioned that they found the arrows hard to see in the Chord diagrams: one person said "The arrows on the Chord diagram sometimes made it harder. The smaller ones were kind of hard to see"; another said "In chord diagram the arrow directions were very difficult to check, specially for thin lines."

6 DISCUSSION

The study found the following main results: participants took longer to answer with Chord diagrams (3.7s per question), with a larger difference for the first of each question type (9.2s); participants also made more errors with Chord (0.1 more errors), again with a larger difference on the first iteration (0.42 more errors); effort scores were higher for Chord diagrams on several scales of the NASA-TLX; and Sankey diagrams were strongly preferred (42 of 51 people).

6.1 Explanations for Results

6.1.1 Completion Time Differences. The completion-time differences seen in our results likely arise from specific visual characteristics of the two types of diagrams. First, the left-to-right arrangement of the Sankey diagram appeared to be easier for participants to comprehend than the circular organization of the Chord diagram, particularly for interpreting link directions and getting an overall sense of the diagram. Several participants made comments about the overall appearance of Chord being disorganized, and several others stated that they found the Sankey diagrams more familiar and understandable. A specific problem involved link direction – with the Sankey diagrams, all links move from left to right, so it is relatively easy to determine the direction of the flow; with Chord, participants needed to spend additional time checking the directions of the links.

Second, the rotated representations may have caused problems for participants (as suggested by previous work, e.g., [18, 52]). Only one participant commented on difficulty reading the rotated text, but the circular layout may have required that participants make additional checks to keep track of of their start position (similar problem to the need for cross-checking sen in [8]).

Third, participants appeared to have more difficulty following links in the Chord diagrams. One problem involved interpreting the link arrowheads. Our visual design used standard arrow parameters from D3, and links were actually wider in Chord diagrams (as described above), but despite the larger size, participants found the arrows difficult to see, indicating a need to emphasize direction information – e.g., by adding a visual marker to arrowheads, or through a second encoding such as moving the origin of outgoing links slightly inward from the circle. Another problem involved the thinning of the ribbons in the centre of the Chord diagrams; when links are narrow (and dynamically change width), it may have been more difficult to trace links from source to destination, and more difficult to estimate link weights.

Fourth, it appeared that counting links was more difficult in the Chord diagrams, particularly in the Debt and Immigration datasets. Problems may have arisen because Chord diagrams group incoming and outgoing links at the same node, meaning that users need to carry out the extra step of visually separating incoming from outgoing links. We note that in the one-way Space dataset (where there was no need to separate incoming from outgoing links), performance of the two types was similar for link-counting tasks.

6.1.2 Differences on the first iteration of a question. Performance with Chord diagrams was worse on the first iteration of each question type, which is an important result because Chord diagrams have often been used in settings where users are unfamiliar with the visualization and will be carrying out analyses for the first time – e.g., mainstream media stories [46] or "Data Storytelling" [7]. The larger difference on the first iteration may be caused by the less-familiar reference frame and layout of Chord diagrams, requiring additional time to understand the overall approach and identify the steps needed to read the visualization. After several iterations, however, participants were able to adapt to the organizational style of Chord diagrams, and there was little performance difference between the two visualizations.

6.2 Generalizing to real-world use, and limitations

Our results are generalizable to real-world use of Chord and Sankey diagrams in several ways. First, our results were consistent across four different datasets and five question types, involving a variety of relationships and analysis topics. All of the datasets had been used in Chord examples in the past, and so were not likely to be biased toward Sankey diagrams. Our question types were adapted from existing taxonomies [2, 40] and were based on realistic explorations of flow-based datasets. Second, our study involved a large number of participants (51), and the MTurk deployment means that our participant sample included a variety of monitor types and sizes, input devices, and computing platforms. Third, we used multiple questions of each type, allowing people to build experience with the two representations.

However, there are also limitations to the study, leading to several opportunities for future work:

- *Duplication of entities in Chord*: Our implementation of Chord diagrams used only one node per entity, and collected both incoming and outgoing links at that node, but it is also possible to duplicate nodes – future studies could test whether separating incoming and outgoing links reduces interpretation problems.
- *Dataset size*: Our datasets were medium-sized (8-15 nodes, 50-100 links), and future work should consider larger datasets. Our hypothesis is that the familiar left-to-right organization of Sankey diagrams will become even more valuable with larger datasets, and that central clutter will increase for Chord diagrams.
- *Connection type*: We tested datasets with directed links, but Chord diagrams are also frequently used for undirected data, and future work should compare the two diagram types on these datasets.
- *Additional question types*: We chose question types that will be common for interpretation tasks with flow visualizations, but future work should expand on this set: e.g., questions that involve a comparison of incoming and outgoing links could see a benefit from Chord's approach of laying out all links in one location (e.g., "Did more users switch *to* Sony or *away from* Sony?").
- *Longer-term use*: Our participants were all initially unfamiliar with either visualization type, and different advantages and disadvantages of the diagrams may emerge with longer-term use and more experienced users.
- *Participants*: MTurk participants may have had little intrinsic interest in learning our particular datasets and tasks, and may have hurried due to the study's fixed payment structure. MTurk provides advantages (i.e., higher ecological validity due to more diversity in participants and viewing environments), but future studies should test groups with specific domain expertise and more visual-analysis experience.
- *Interactivity*: Our diagrams did not include interaction (e.g., zooming or hover highlighting), in order to better assess visual interpretability. However, interactivity provides additional capabilities that could overcome visual limitations, and future studies should also test these features.
- *Enhancements to visualization designs*: The use of standard visualization parameters available in D3 [5] limited us in terms of adding potential enhancements to our diagram types. For example, researchers have proposed layout metrics that could reduce clutter in both Chord and Sankey diagrams [12, 37]. Future studies should explore whether enhancements could change the interpretability and usability of both diagram types.

7 CONCLUSION

Visualizing connections and flows can be accomplished using several approaches, with Chord diagrams (using radial layout) and Sankey diagrams (using linear layout) both popular. Previous work suggests both advantages and disadvantages of the radial approach, but little is known about the usability and interpretability of these two layout styles for flow visualizations. We carried out a crowdsourced study to compare Chord and Sankey diagrams, and our results provide designers with a better understanding of their performance and usability, as well as adding a further comparison of radial and linear organizations of data. Participants took substantially longer to answer questions with Chord diagrams and also made more errors; they also rated Chord as requiring more effort, and strongly preferred Sankey diagrams – a main reason was that Sankey's left-to-right organization provided a familiar frame of reference for interpreting the diagram. Our study identifies and explains limitations of radial Chord diagrams, and shows that some of the problems with radial layouts seen in previous work also affect visualizations of flow datasets.

ACKNOWLEDGMENTS

Thanks to our study participants, and to our reviewers for their feedback. This research was supported by the Natural Sciences and Engineering Research Council of Canada, and the Plant Phenotyping and Imaging Research Centre.

REFERENCES

[1] Yael Albo, Joel Lanir, Peter Bak, and Sheizaf Rafaeli. 2015. Off the radar: Comparative evaluation of radial visualization solutions for composite indicators. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 569–578.

[2] Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 111–117.

[3] Ruth Bollongino, Olaf Nehlich, Michael P Richards, Jörg Orschiedt, Mark G Thomas, Christian Sell, Zuzana Fajkošová, Adam Powell, and Joachim Burger. 2013. 2000 years of parallel societies in Stone Age Central Europe. *Science* 342, 6157 (2013), 479–481.

[4] Mike Bostock. 2020. *Directed Chord Diagram.* https://observablehq.com/@d3/directed-chord-diagram

[5] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2301–2309.

[6] Matthew Brehmer, Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe. 2018. Visualizing ranges over time on mobile phones: a task-based crowdsourced evaluation. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 619–629.

[7] Nadieh Bremer. 2021. Using Data Storytelling with a chord diagram. https://www.visualcinnamon.com/2014/12/using-data-storytelling-with-chord/

[8] Michael Burch, Fabian Beck, and Daniel Weiskopf. 2012. Radial Edge Splatting for Visualizing Dynamic Directed Graphs.. In *GRAPP/IVAPP*. 603–612.

[9] Michael Burch, Felix Bott, Fabian Beck, and Stephan Diehl. 2008. Cartesian vs. radial–a comparative evaluation of two visualization tools. In *International Symposium on Visual Computing*. Springer, 151–160.

[10] Michael Burch and Daniel Weiskopf. 2014. On the benefits and drawbacks of radial diagrams. In *Handbook of human centric visualization*. Springer, 429–451.

[11] William S Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387 (1984), 531–554.

[12] Frederik L Dennig, Maximilian T Fischer, Michael Blumenschein, Johannes Fuchs, Daniel A Keim, and Evanthia Dimara. 2021. ParSetgnostics: Quality Metrics for Parallel Sets. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 375–386.

[13] Sara Di Bartolomeo, Aditeya Pandey, Aristotelis Leventidis, David Saffo, Uzma Haque Syeda, Elin Carstensdottir, Magy Seif El-Nasr, Michelle A Borkin, and Cody Dunne. 2020. Evaluating the effect of timeline shape on visualization task performance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[14] Stephan Diehl, Fabian Beck, and Michael Burch. 2010. Uncovering strengths and weaknesses of radial visualizations—an empirical approach. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 935–942.

[15] Diego Doval, Spiros Mancoridis, and Brian S Mitchell. 1999. Automatic clustering of software systems using a genetic algorithm. In *STEP'99. Proceedings Ninth International Workshop Software Technology and Engineering Practice*. IEEE, 73–81.

[16] Geoffrey M Draper, Yarden Livnat, and Richard F Riesenfeld. 2009. A survey of radial methods for information visualization. *IEEE transactions on visualization and computer graphics* 15, 5 (2009), 759–776.

[17] Alberto Ferrari and Marco Russo. 2016. *Introducing Microsoft Power BI*. Microsoft Press.

[18] Douglas J Gillan. 2000. A Componential Model of Human Interaction with Graphs. V: Using Pie Graphs to Make Comparisons. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 44. SAGE Publications Sage CA: Los Angeles, CA, 3–439.

[19] Joseph Goldberg and Jonathan Helfman. 2011. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information visualization* 10, 3 (2011), 182–195.

[20] Novi Handayani. 2021. Chord Diagram With Immigration Data. https://observablehq.com/@novi05211640000015/chord-diagram

[21] Christopher Healey and James Enns. 2011. Attention and visual memory in visualization and computer graphics. *IEEE transactions on visualization and computer graphics* 18, 7 (2011), 1170–1188.

[22] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *Proceedings of the 28th Annual Chi Conference on Human Factors in Computing Systems* (2010), 203–212. https://doi.org/10.1145/1753326.1753357

[23] Haoyan Huo, Ziqin Rong, Olga Kononova, Wenhao Sun, Tiago Botari, Tanjin He, Vahe Tshitoyan, and Gerbrand Ceder. 2019. Semi-supervised machine-learning classification of materials synthesis procedures. *Npj Computational Materials* 5, 1 (2019), 1–7.

[24] Ross Ihaka and Robert Gentleman. 1996. R: a language for data analysis and graphics. *Journal of computational and graphical statistics* 5, 3 (1996), 299–314.

[25] Mikel Iturbe, Inaki Garitano, Urko Zurutuza, and Roberto Uribeetxeberria. 2016. Visualizing Network Flows and Related Anomalies in Industrial Networks using Chord Diagrams and Whitelisting.. In *VISIGRAPP (2: IVAPP)*. 101–108.

[26] Amin Jalali. 2016. Reflections on the use of chord diagrams in social network visualization in process mining. In *2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 1–6.

[27] Richard Johnsonbaugh and Martin Kalin. 1991. A graph generation software package. In *Proceedings of the twenty-second SIGCSE technical symposium on Computer Science Education*. 151–154.

[28] Pierre Jolicoeur, Shimon Ullman, and Marilynn Mackay. 1986. Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition* 14, 2 (1986), 129–140.

[29] Kevin Jolly. 2018. *Hands-on data visualization with Bokeh: Interactive web plotting for Python using Bokeh*. Packt Publishing Ltd.

[30] Stephen G Kobourov, Sergey Pupyrev, and Bahador Saket. 2014. Are crossings important for drawing large graphs?. In *International Symposium on Graph Drawing*. Springer, 234–245.

[31] Asher Koriat and Joel Norman. 1985. Reading rotated words. *Journal of Experimental Psychology: Human Perception and Performance* 11, 4 (1985), 490.

[32] Martin Krzywinski, Jacqueline Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J Jones, and Marco A Marra. 2009. Circos: an information aesthetic for comparative genomics. *Genome research* 19, 9 (2009), 1639–1645.

[33] Kevin C. Lam, Carl Gutwin, Madison Klarkowski, and Andy Cockburn. 2021. The Effects of System Interpretation Errors on Learning New Input Mechanisms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 713, 13 pages. https://doi.org/10.1145/3411764.3445366

[34] Timothy R Levine and Craig R Hullett. 2002. Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research* 28, 4 (2002), 612–625.

[35] Miriah Meyer, Tamara Munzner, and Hanspeter Pfister. 2009. MizBee: a multiscale synteny browser. *IEEE transactions on visualization and computer graphics* 15, 6 (2009), 897–904.

[36] Tamara Munzner, Eric Hoffman, K Claffy, and Bill Fenner. 1996. Visualizing the global topology of the MBone. In *Proceedings IEEE Symposium on Information Visualization'96*. IEEE, 85–92.

[37] Jorge Dionisio Nunez Siri. 2019. *AccuSyn: Using Simulated Annealing to Declutter Genome Visualizations*. Ph. D. Dissertation. University of Saskatchewan.

[38] Stephen Olejnik and James Algina. 2003. Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological methods* 8, 4 (2003), 434.

[39] Wei Qi, Guy J Abel, Raya Muttarak, and Shenghe Liu. 2017. Circular visualization of China's internal migration flows 2010–2015. *Environment and Planning A: Economy and Space* 49, 11 (2017), 2432–2436. https://doi.org/10.1177/0308518X17718375 arXiv:https://doi.org/10.1177/0308518X17718375

[40] Ghulam Jilani Quadri and Paul Rosen. 2021. A survey of perception-based visualization studies by task. *IEEE Transactions on Visualization and Computer Graphics* (2021).

[41] Jessica Saben, Ying Zhong, Samantha McKelvey, Nafisa K Dajani, Aline Andres, Thomas M Badger, Horacio Gomez-Acevedo, and Kartik Shankar. 2014. A comprehensive analysis of the human placenta transcriptome. *Placenta* 35, 2 (2014), 125–131.

[42] Lauren M Scott and Mark V Janikas. 2010. Spatial statistics in ArcGIS. In *Handbook of applied spatial analysis*. Springer, 27–41.

[43] JN Siri, Eric Neufeld, Isobel Parkin, and Andrew Sharpe. 2020. Using Simulated Annealing to Declutter Genome Visualizations. In *The Thirty-Third International Flairs Conference*.

[44] Stephen Smart, Keke Wu, and Danielle Albers Szafir. 2020. Color Crafting: Automating the Construction of Designer Quality Color Ramps. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 1215–1225. https://doi.org/10.1109/TVCG.2019.2934284

[45] Danielle Albers Szafir, Maureen Stone, and Michael Gleicher. 2014. Adapting color difference for design. In *Color and Imaging Conference*, Vol. 2014. Society for Imaging Science and Technology, 228–233.

[46] The New York Times. 2007. Close-Ups of the Genome, Species by Species by Species. https://archive.nytimes.com/www.nytimes.com/imagepages/2007/01/22/science/20070123_SCI_ILLO.html

[47] Miles A Tinker. 1956. Effects of angular alignment upon readability of print. *Journal of Educational Psychology* 47, 6 (1956), 358.

[48] Urmila. 2021. Space Tech Companies and Venture Capitals. https://observablehq.com/@urmilaj/space-tech-companies-and-venture-capitals

[49] Manuela Waldner, Alexandra Diehl, Denis Gračanin, Rainer Splechtna, Claudio Delrieux, and Krešimir Matković. 2019. A comparison of radial and linear charts for visualizing daily patterns. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 1033–1042.

[50] Colin Ware, Helen Purchase, Linda Colpoys, and Matthew McGill. 2002. Cognitive measurements of graph aesthetics. *Information visualization* 1, 2 (2002), 103–110.

[51] Marc Weber, Marc Alexa, and Wolfgang Müller. 2001. Visualizing time-series on spirals.. In *Infovis*, Vol. 1. 7–14.

[52] Daniel Wigdor and Ravin Balakrishnan. 2005. Empirical investigation into the effect of orientation on text readability in tabletop displays. In *ECSCW 2005*. Springer, 205–224.

[53] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[54] Linda Woodburn, Yalong Yang, and Kim Marriott. 2019. Interactive visualisation of hierarchical quantitative data: an evaluation. In *2019 IEEE Visualization Conference (VIS)*. IEEE, 96–100.