




The Manhattan Wheel: A Radial Visualization Story for Genome-wide Association Study Data

K. E. Zhang¹ , M. Vaudel^{1,2} , and L. A. Garrison¹ 

¹Dept. of Informatics, Univ. of Bergen, Norway ² Dept. of Clinical Science, Mohn Center for Diabetes Precision Medicine, Univ. of Bergen, Norway

Abstract

Genome-wide association studies (GWAS) are critical to identifying genetic variations associated with a particular trait or disease. It is important to cultivate an awareness of GWAS in the general public as members of this group are key participants of these studies. However, low genetic data literacy and trust in the sharing of genetic data pose challenges to learning and engaging with GWAS concepts. In this design study, we explore design strategies for the public communication of GWAS data. As part of this study, we present an interactive visual prototype that explores the use of narrative structure, linked visualizations through scrollytelling, and plain language to onboard and communicate genetic concepts to a GWAS-naive audience.

CCS Concepts

• **Human-centered computing** → **Information visualization**; • **Applied computing** → **Genomics**;

1. Introduction

Genome-wide association studies (GWAS) compare entire genomes in a population to identify genetic variations that are associated with a particular phenotypic, i.e., observable, trait or disease [UHM*21]. In humans, single nucleotide polymorphisms (SNPs) are the most common type of genetic variation. Domain experts routinely identify and compare significant SNPs by visualizing GWAS data using a Manhattan plot—a linear scatter plot displaying dense data. Millions of SNPs, ordered by genomic position and grouped by chromosome, are plotted against $-\log_{10}$ transformed P values. It is important to cultivate an awareness of GWAS in the general public as they are major participants and beneficiaries of these studies. The success of contemporary biomedical and public health research depends on public contribution of genetic and other biological samples to biobanks (e.g., UK Biobank [OSP05]), which researchers then access to study the genes, environment, lifestyle, and other mechanisms associated with disease in order to improve public health outcomes. Multiple challenges factor into the public communication of GWAS, including low genetic literacy [MMA*20], variable statistical literacy [RHS20], scientific uncertainty, and poor public perception of genetic research and genetic data sharing [MMA*20].

Thus far, few works in the public health space integrate visual analytics and public dissemination [PL20], and visualizations for a GWAS-naive audience have not been well-characterized. Design strategies that increase comprehension, interest, and user engagement may help facilitate the communication of science to a public audience [Dah14]. Notable examples include scrollytelling [SZ18] websites from FinnGen [Aal23] and the Broad Institute [Bro21]

that convey the importance of genetic data using aesthetically-pleasing visualizations and jargon-free explanations. To further explore these strategies, this poster presents an interactive visual prototype that integrates narrative visualization [SH10], plain language [AWB*23], radial visualization [KSB*09], scrollytelling, and onboarding strategies [SGP*19] to communicate GWAS data on child growth to a GWAS-naive audience (Fig. 1).

2. Data & Task Abstraction

We derived a set of user tasks and designs informed by literature and consultation with four domain experts, one of whom is a co-author. The final prototype was created through a process of iterative sketching using the five design-sheet methodology [RHR15]. We used Brehmer & Munzner’s task typology [BM13] with principles in perception and engagement to justify our design choices.

Data Processing. We used GWAS summary data on early childhood body mass index (BMI) [HVSN*22] from the Norwegian Mother, Father and Child Cohort Study (MoBa) [MBV*16]. The complete dataset includes 12 collected time points over several years and are mixed-type (qualitative and quantitative). For this prototype, we used a subset of data, namely rsID (SNP identifier), chromosome, genomic position, effect size, standard error and P value of SNPs that either A) were not spurious correlations at any of the 12 time points (P value of $1e-4$ or lower), or B) were one of 40 SNPs of research interest identified in Helgeland et al. [HVSN*22].

Task Abstraction. Inspired by role-playing games for public engagement in scientific citizenship [GS12], we aimed to increase awareness of the purpose and relevancy of GWAS by walking a

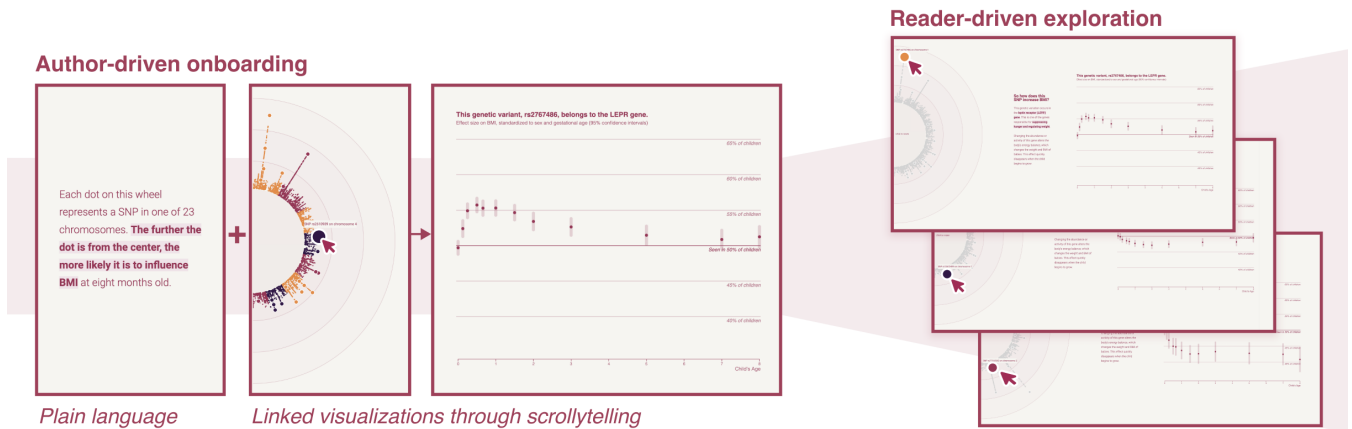


Figure 1: Our approach to visually communicating genome-wide association study (GWAS) data uses a martini-glass narrative structure, linked visualizations through scrollytelling, and plain language to onboard and communicate GWAS concepts to a GWAS-naive audience.

GWAS-naive user through the tasks of a genetics researcher (domain expert). We designed versions of domain expert tasks where the target user needed to gain a gist of the data but not report on specific, accurate values. The first task is to identify significant SNPs (discover outliers) in the Manhattan plot (**T1**). Once these SNPs are identified, one looks at the overall effect (discover trends) of a SNP on childhood BMI in a separate dot plot (**T2**). An analyst will also want to compare the effects of different SNPs, but we excluded this task on reflection to avoid overloading a GWAS-naive user.

2.1. The Manhattan Wheel & Next Steps

The visualization story follows a martini-glass narrative structure [SH10] that leads the user through two linked visualizations: a semi-hidden radial Manhattan plot on the left side of the screen and a linear dot plot on the right. In the center of the two visualizations lies the narrative text. The website starts with a scroll-driven narrative that onboards the user to interpreting GWAS visualizations and ends with free user exploration of the two visualizations. The application is open-source and available at <https://github.com/amykzhang/manhattan-wheel>.

Visual Encodings. We used a radial representation of the Manhattan plot to 1) employ the memorable and aesthetically-pleasing nature of radial layouts for user engagement [BVB*13,SMYL24] and 2) exploit radial distortion to push outliers further away from the center of the plot (**T1**). We used opacity and size to make salient either SNPs of interest (**T1**) or a user-selected SNP. For color, we followed genomic visualization conventions which use an alternating color palette. We chose a warm color scheme, adjusted for contrast and web accessibility, to elicit excitement and playfulness [BPS17].

Chart Elements & Text. Onboarding text was written in plain language (Flesch-Kincaid readability score of 60-70 [Fle79]) and its scientific accuracy was checked by domain experts. This text aimed to teach users how to identify outliers in the Manhattan plot (**T1**) and interpret trends in the dot plot (**T2**). We revised chart elements that were difficult to grasp for audiences with low visualization lit-

eracy e.g., addition of chart annotations and omission of axis ticks related to statistics (i.e., P and beta values).

Interactions & User Flow. We followed Shneiderman’s visual information-seeking mantra [Shn96] for both the onboarding and free exploration experience. The user first sees a Manhattan plot with fully salient data points. When a user selects a SNP on the Manhattan plot, additional information on the SNP’s effect size on early childhood BMI is revealed on the dot plot (**T2**). For both plots, hovering over data points reveals their individual details on demand. To promote engagement with the visualization, we hid half of the Manhattan plot offscreen to encourage users to rotate and explore the full extent of the Manhattan plot.

Reflections & Next Steps. In this work, we explored design strategies that may not necessarily follow perceptual design guidelines for visual data communication. Aesthetic, memorable, and whimsical radial visualizations may provide benefits to a non-expert audience that offsets its disadvantages. In non-expert GWAS data communication, radial distortion—considered a trade-off of radial layouts—may even support the user in accomplishing tasks such as the identification of outliers. In creating bespoke visualizations with large, multidimensional datasets, we found it challenging to make the visualizations accessible e.g., via a screen-reader. Including plain language text, chart annotations, and omitting jargon-filled elements may make the data more inclusive for a broader audience [LCI*20]. Finally, we plan to conduct an end-user evaluation that compares conventional and unconventional design strategies for communication of GWAS concepts to naive audiences.

Acknowledgements

The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. This work is funded by Univ. of Bergen and Trond Mohn Foundation (#813558, Visualizing Data Science for Large Scale Hypothesis Management in Imaging Biomarker Discovery).

References

- [Aal23] AALTO UNIVERSITY, DEPARTMENT OF ART AND MEDIA: Can one person living in finland contribute to the future of global medicine?, 2023. URL: <https://geneviz.aalto.fi/what-is-finngen/about.html>. 1
- [AWB*23] AUGUST T., WANG L. L., BRAGG J., HEARST M. A., HEAD A., LO K.: Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–38. doi:10.1145/3589955. 1
- [BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2376–2385. doi:10.1109/TVCG.2013.124. 1
- [BPS17] BARTRAM L., PATRA A., STONE M.: Affective color in visualization. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 1364–1374. doi:10.1145/3025453.3026041. 2
- [Bro21] BROAD INSTITUTE: Polygenic scores, 2021. URL: <https://polygeniccores.org/explained/>. 1
- [BVB*13] BORKIN M. A., VO A. A., BYLINSKII Z., ISOLA P., SUNKAVALLI S., OLIVA A., PFISTER H.: What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315. doi:10.1109/TVCG.2013.234. 2
- [Dah14] DAHLSTROM M. F.: Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences* 111, supplement_4 (2014), 13614–13620. doi:10.1073/pnas.1320645111. 1
- [Fle79] FLESCHE R.: *How to write plain English: A book for lawyers and consumers*, vol. 76026225. Harper & Row New York, 1979. 2
- [GS12] GAYDOS M. J., SQUIRE K. D.: Role playing games for scientific citizenship. *Cultural Studies of Science Education* 7 (2012), 821–844. doi:10.1007/s11422-012-9414-2. 1
- [HVS*22] HELGELAND Ø., VAUDEL M., SOLE-NAVAIS P., FLATLEY C., JUODAKIS J., BACELIS J., KOLØEN I. L., KNUDSEN G. P., JOHANSSON B. B., MAGNUS P., ET AL.: Characterization of the genetic architecture of infant and early childhood body mass index. *Nature Metabolism* 4, 3 (2022), 344–358. doi:10.1038/s42255-022-00549-1. 1
- [KSB*09] KRZYWINSKI M., SCHEIN J., BIROL I., CONNORS J., GASCOYNE R., HORSMAN D., JONES S. J., MARRA M. A.: Circos: an information aesthetic for comparative genomics. *Genome Research* 19, 9 (2009), 1639–1645. doi:10.1101/gr.092759.109. 1
- [LCI*20] LEE B., CHOE E. K., ISENBERG P., MARRIOTT K., STASKO J.: Reaching broader audiences with data visualization. *IEEE Computer Graphics and Applications* 40, 2 (2020), 82–90. doi:10.1109/MCG.2020.2968244. 2
- [MBV*16] MAGNUS P., BIRKE C., VEJRUP K., HAUGAN A., ALSAKER E., DALTVEIT A. K., HANDAL M., HAUGEN M., HØISETH G., KNUDSEN G. P., ET AL.: Cohort profile update: the norwegian mother and child cohort study (moba). *International Journal of Epidemiology* 45, 2 (2016), 382–388. doi:10.1093/ije/dyw029. 1
- [MMA*20] MIDDLETON A., MILNE R., ALMARRI M. A., ANWER S., ATUTORNU J., BARANOVA E. E., BEVAN P., CEREZO M., CONG Y., CRITCHLEY C., ET AL.: Global public perceptions of genomic data sharing: what shapes the willingness to donate dna and health data? *The American Journal of Human Genetics* 107, 4 (2020), 743–752. doi:10.1016/j.ajhg.2020.08.023. 1
- [OSP05] OLLIER W., SPROSEN T., PEAKMAN T.: Uk biobank: From concept to reality. *Pharmacogenomics* 6, 6 (2005), 639–646. doi:10.2217/14622416.6.6.639. 1
- [PL20] PREIM B., LAWONN K.: A survey of visual analytics for public health. *Computer Graphics Forum* 39, 1 (2020), 543–580. doi:10.1111/cgf.13891. 1
- [RHR15] ROBERTS J. C., HEADLEAND C., RITSOS P. D.: Sketching designs using the five design-sheet methodology. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 419–428. doi:10.1109/TVCG.2015.2467271. 1
- [RHS20] RUNGE J., HUDSON-SHARP N.: *Public Understanding of Economics and Economic Statistics*. Economic Statistics Centre of Excellence Occasional Papers ESCOE-OP-03, Economic Statistics Centre of Excellence, Nov 2020. URL: <https://ideas.repec.org/p/nsr/escoeo/escoe-op-03.html>. 1
- [SGP*19] STOIBER C., GRASSINGER F., POHL M., STITZ H., STREIT M., AIGNER W.: Visualization onboarding: Learning how to read and use visualizations, Aug 2019. doi:10.31219/osf.io/c38ab. 1
- [SH10] SEGEL E., HEER J.: Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148. doi:10.1109/TVCG.2010.179. 1, 2
- [Shn96] SHNEIDERMAN B.: The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages* (1996), pp. 336–343. doi:10.1109/VL.1996.545307. 2
- [SMYL24] STÅHLBOM E., MOLIN J., YNNERMAN A., LUNDSTRÖM C.: Should i make it round? suitability of circular and linear layouts for comparative tasks with matrix and connective data. *Computer Graphics Forum* 43, 3 (2024), e15102. doi:10.1111/cgf.15102. 2
- [SZ18] SEYSER D., ZEILLER M.: Scrollytelling—an analysis of visual storytelling in online journalism. In *2018 22nd International Conference Information Visualisation (IV'18)* (2018), IEEE, pp. 401–406. doi:10.1109/iv.2018.00075. 1
- [UHM*21] UFFELMANN E., HUANG Q. Q., MUNUNG N. S., DE VRIES J., OKADA Y., MARTIN A. R., MARTIN H. C., LAPPALAINEN T., POSTHUMA D.: Genome-wide association studies. *Nature Reviews Methods Primers* 1, 1 (2021), 59. doi:10.1038/s43586-021-00056-9. 1