The Design of "X by Y"

An Information-Aesthetic Exploration of the Ars Electronica Archives

Moritz Stefaner

THIS CHAPTER PRESENTS THE PROJECT "X BY Y," a visualization of all entries to the Prix Ars Electronica, a well-known media art award, from 1987 to 2009. The final version of the visualization consists of a series of large-scale prints, splitting up the submissions according to different criteria. This chapter describes the process leading up to the final piece, and the rationale for specific design decisions.

Briefing and Conceptual Directions

The Ludwig Boltzmann Institute for media.art.research contacted me in spring 2009 to work on the submission databases for the Prix Ars Electronica. The media art festival Ars Electronica had its 30th anniversary that year, and together, we decided to take on the challenge of trying to visually analyze all the submissions to the Prix over its 22-year history. The databases containing the submission information had never before been analyzed in their entirety.

In the kickoff meeting for the project, we discussed our objectives. The creative lead of the whole visualization project, Dietmar Offenhuber, explained that different visualizations were to be developed in order to study three different angles on the festival's history:

Quantitative analysis
　　What can we say about the festival by looking at the submissions over the years? How do the various categories differ, where do the submissions come from, and how do the values change over time?

Social networks

> Who were the jury members throughout the years? How are they—as well as the awarded artists—connected to one another?

Art historical context

> What impact have the awarded projects had? Where have they been referenced, and how have they influenced the field of media arts?

The project I was to work on would belong to the first category. Specifically, I was to investigate what hypotheses and insights we could generate by looking into the submission data, and whether we could find an appropriate visual to convey the characteristics of the "ars world" to visitors of the exhibition.

Together with the art historians working on the Ars Electronica archives, I tried to define some first directions of interest, reflected in the matrix in Figure 13-1. Without looking at the databases in detail, it was assumed that we should be able to work on basic dimensions like the submission's author, country, year, prize category, and keywords, as well as whether it received a prize. The matrix reflects the *a priori* interest in certain combinations of these factors—i.e., where the experts expected interesting findings to emerge. For instance, it was assumed that we might want to split winners by country (and compare this data to the overall submission statistics) and look at the relationships between authors and categories.

	Author	Country	Year	Category	Keywords	Winner?
Author			✖	✖	✖	✖
Country			✖	✖	✖	✖
Year					✖	
Category					✖	
Keywords						✖
Winner?						

Figure 13-1. *Matrix of initial interest in attribute combinations*

Understanding the Data Situation

Next, I began to look into the available data, together with Sandor Herramhof. Over the years, a number of database schemas with different conventions and varying degrees of modeling detail had been used, which made it very difficult to get an early overview of the existing data. For instance, one database featured additional information stored in an XML format inside a text field, but only for some of the submissions. In order to facilitate the process of acquiring an overview of the data, I developed *dbcounter*,[*] a small nodebox script that would enable us to quickly get an overview of large sets of categorical data. *dbcounter* walks over a CSV file, determines all the unique value attributes, counts how often they occur, and plots the output as an area chart. The gray areas (see Figure 13-2) indicate missing or NULL values. Overall, the tool proved useful for understanding the contents of our databases, especially in finding missing values and getting an idea of the data diversity.
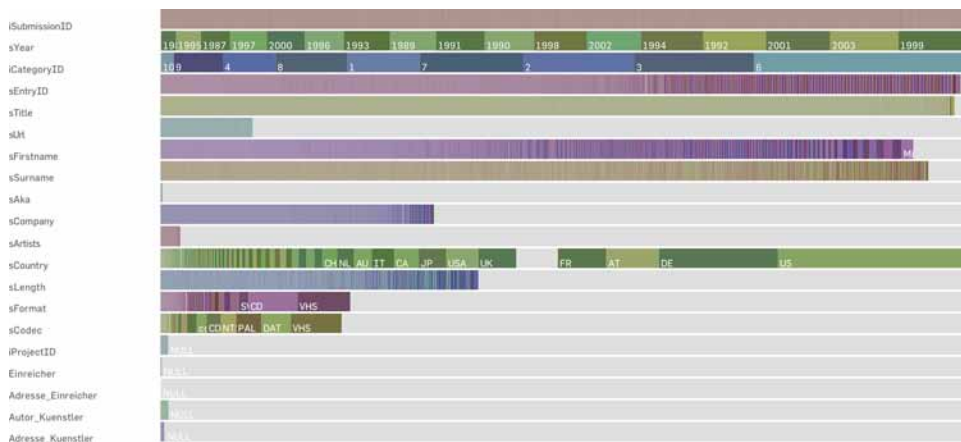

Figure 13-2. *First overview of the database contents with dbcounter, a custom nodebox script*

From these plots, some facts about the databases quickly became clear:

- There were a number of apparently redundant fields, such as "Land" (German for "country") and "sYear," caused by the merging of database schemas over the years.

- Names, years, and categories were present fairly completely.

- Much less country, company, and web address information was present than expected.

[*] See *http://well-formed-data.net/archives/306/dbcounter-quick-visual-database-stats*.

On the one hand, this quick first analysis allowed us to understand what types of attribute combinations could be expected to be meaningful and to cover at least a large part of the data. As the database migration was an ongoing process, it also provided us with a useful overview of the areas in which we should seek to improve the data, and which fields could be combined or filled up more completely. For instance, the team working on the databases containing the representation country field was in fact trying to complete as much of the information as possible ("It seems like really interesting information, and we are already almost there").

Exploring the Data

After the first quantitative analysis of the individual fields, the next step was to slice and dice a preliminary subset of the data, to investigate correlations and get some hints about the reasons for some of the gaps in the data. For this step we used the commercial software Tableau,* which allowed us to explore the data in the spreadsheets we imported and the databases we connected to using interactive charts in a flexible and expressive workspace. For instance, we used Tableau to characterize the submissions missing country information by year and category (see Figure 13-3), in order to identify the biggest gaps and facilitate the search for the missing information in other media, such as catalog texts. Questions like "How does the number of submissions relate to the submission categories?" and "Has this changed over the years?" can also be answered quite easily in a graphical user interface.

Other explorations included a characterization of companies in terms of the categories in which they had submitted entries. The chart in Figure 13-4, for instance, revealed the potential for some interesting stories. However, it also quickly became clear that a large amount of manual work would be required to clean up all the variations in the spellings of the company names in the different databases if we wanted to be able to make accurate statements.

We also used Tableau to produce an initial world map of the submissions (see Figure 13-5), with the pie charts for each country indicating the category distributions. This early map reveals the European/U.S.-centric nature of the festival. It soon became clear that this simplistic approach to producing a cartogram would be inefficient for this skewed data distribution, motivating the more elaborate approach presented later.
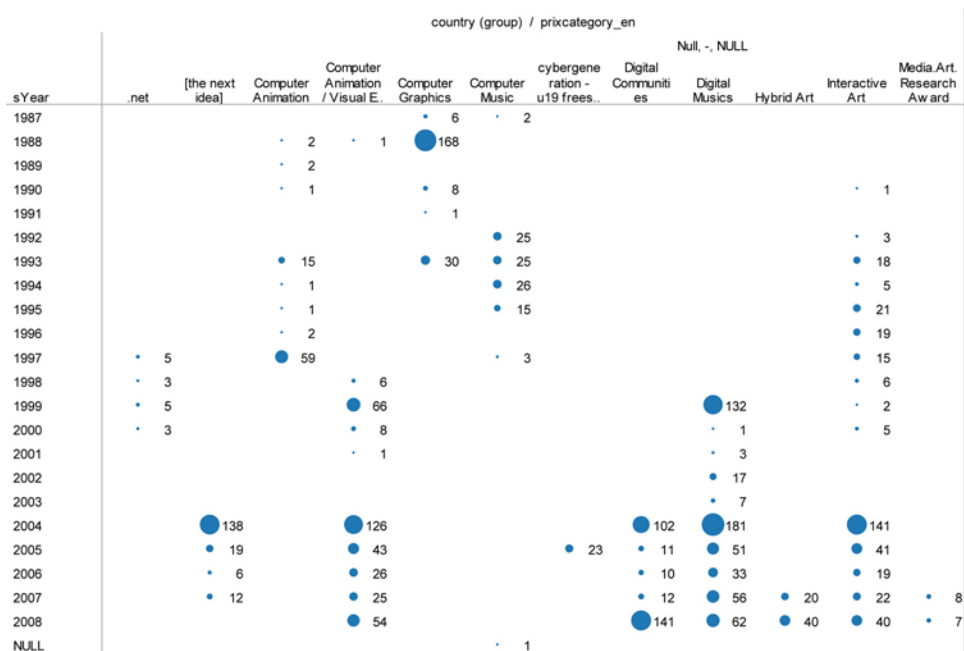
* See *http://www.tableausoftware.com*.

country (group) / prixcategory_en

sYear	.net	[the next idea]	Computer Animation	Computer Animation / Visual E.	Computer Graphics	Computer Music	cybergeneration - u19 frees..	Digital Communities	Digital Musics	Hybrid Art	Interactive Art	Media.Art. Research Award
1987					6	2						
1988			2	1	168							
1989			2									
1990			1		8						1	
1991					1							
1992						25					3	
1993			15		30	25					18	
1994			1			26					5	
1995			1			15					21	
1996			2								19	
1997	5		59			3					15	
1998	3			6							6	
1999	5			66					132		2	
2000	3			8					1		5	
2001				1					3			
2002									17			
2003									7			
2004		138		126				102	181		141	
2005		19		43			23	11	51		41	
2006		6		26				10	33		19	
2007		12		25				12	56	20	22	8
2008				54				141	62	40	40	7
NULL						1						

Figure 13-3. *A plot of submissions with missing country information, split up by year and category*
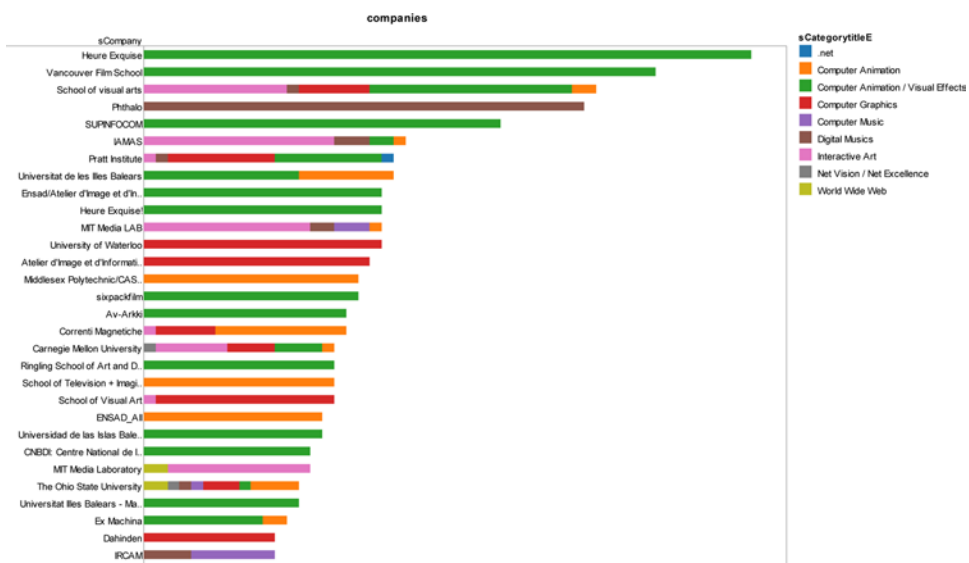
Figure 13-4. *Submissions by company or institution, colored by categories*

Figure 13-5. *World map with submissions per country, split up by category*

I also explored some of the data in Microsoft Excel, as it seemed superior at producing stacked charts we could use for investigating trends over the years or comparing attribute distributions in subsets of the data. For instance, Figure 13-6 shows the relative proportion of submissions and the different types of prizes won by each country. From this chart, it appeared as if the U.S. were responsible for about 30% of all submissions but won over 60% of all Golden Nicas (the highest prize awarded). However, this trend turned out to be much less pronounced in the full and verified set of data analyzed later. We were also aware that the relation of countries to prizes won is a complex and sensitive matter that can only be fully understood by considering various other aspects of the data, such as the number of submissions in each category (for instance, the computer graphics categories in the 1980s had staggering numbers of submissions compared to other categories). So, while there was potential for some interesting insights, we decided that we would present this story only if we were able to provide some context and explanation.
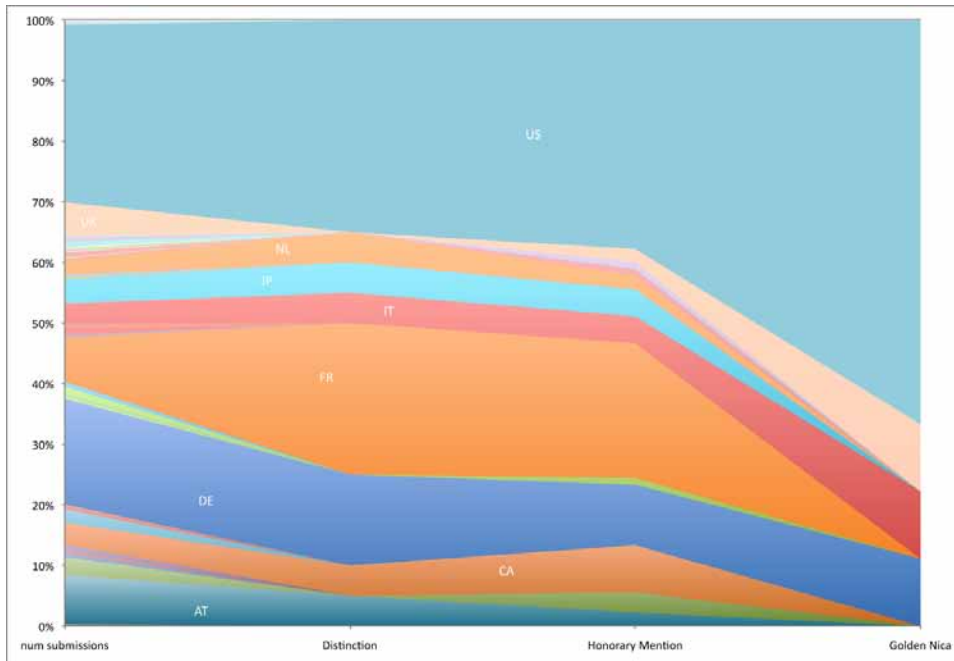
Figure 13-6. *Prizes won by different countries*

First Visual Drafts

The analytic process delivered some initial insights into the data, and gave my collaborators enough opportunities—maybe more than they desired—to correct, clean, and complete the databases. On that basis, borrowing terminology from Tom Armitage's BERG blog post "Toiling in the data-mines: What data exploration feels like,"* I had a good sense of what was *available*, *significant*, and *interesting*, and of the *scale* of the data. The next step was to work on the visualization principles.

To quickly prototype some different visual options, I switched to Flash ActionScript 3 using the *flare* library,† an advanced general-purpose framework for producing interactive visualizations, and I explored more of the stacked charting options using the Excel charts I started with. One insight I gleaned from these charts was that we should try harder to emphasize the individual data points (e.g., the individual years on the vertical axis in Figure 13-7), rather than producing continuous stacked area charts. In the Ars Electronica case, submissions are made on an annual basis only, so a visual interpolation between years would have been misleading and a distortion of reality.

* See *http://berglondon.com/blog/2009/10/23/toiling-in-the-data-mines-what-data-exploration-feels-like/*.
† See *http://flare.prefuse.org*.

These considerations led to the development of more fragile charts, with the inter-
polation areas toned down to support the notion of them being only connectors
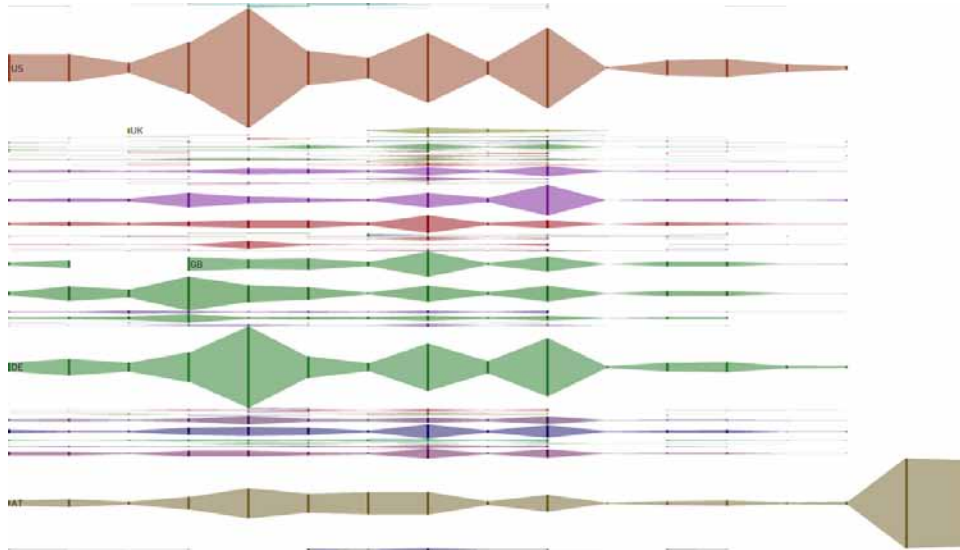between more "solid" yearly events.

Figure 13-7. *A first attempt at displaying categories by country*

Exploring stacked area charts for categories over the years revealed some addi-
tional issues to tackle from a conceptual point of view. The category structure of
Ars Electronica underwent a continuous evolution over the years. For instance, the
"Computer Music" category was not present in 1991, yet it was in the years before
and after. Then, in 1999, it was discontinued and a new category, "Digital Musics,"
was added. How best to treat this situation is a tricky conceptual question: on the one
hand, these are clearly related categories, but on the other hand, it might be too sim-
plistic to unify them and treat them as the same category with a different label. For
decisions like these, expert opinions and the designer's view have to be taken into
account to formulate an accurate, yet pragmatic and understandable, approach. After
some discussions, we resolved the issue by treating these as independent categories but
giving them identical colors in the different visualizations (Figure 13-8).
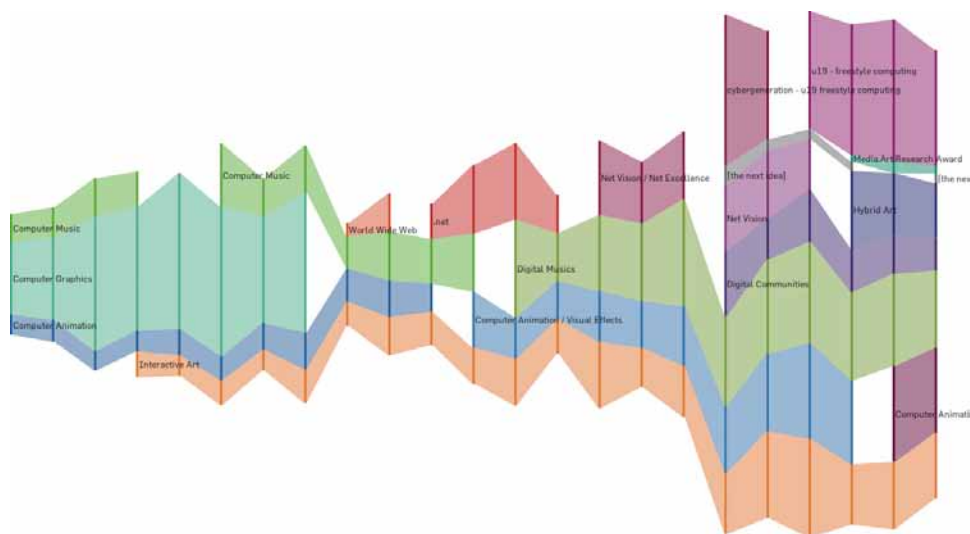
Figure 13-8. *Categories over the years*

I also became more interested in the evocative, implicit communication aspects of the visualization as I explored the existing charts. I felt uncomfortable with their character; from a visual point of view the Flare charts looked appealing, yet a bit too fragile. However, there was also a much bigger concern: while it is interesting to approach a cultural phenomenon like a media art prize in purely quantitative terms, we felt as though we were losing a sense of the scale and diversity of the data and characterizing it in strokes that were too broad. Effective visualization has a strong relation to summarization and prioritization; however, simply creating some rather abstract charts would not have done the topic itself justice. Might not there be a way to display the totals, fractions, and interrelationships without neglecting or even hiding the individual submissions?

The Visual Principle

This motivation led me first to explore dense pixel mosaic displays (Keim 2000), following the idea that I would like to see one visual marker for each individual submission. To get a sense of how many points I could fit on a standard screen, I did some quick tests using random data (see Figure 13-9).
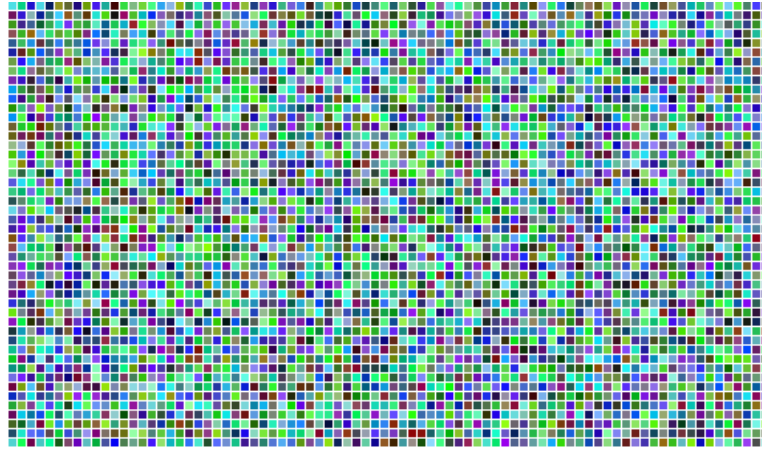
Figure 13-9. *Experimenting with dense pixel displays*

I found the results quite encouraging and decided to investigate further by looking at *QR codes*.* Could we actually build QR codes with meaningful URLs that also worked as area- or pixel-based data graphics? Another idea was to do something along the lines of Wattenberg's (2005) colored segments of space-filling curves to produce diagrams similar to treemaps (so-called "jigsaw maps").

The real eureka moment, however, came when I remembered a placement algorithm I had used in an earlier project. Computed on the basis of the *golden angle* (the angle corresponding to a "golden section" of a full circle, or 137.5 degrees), it imitates the arrangement of sunflower seeds—the most efficient and visually mesmerizing way of packing small elements into a large circle. Figure 13-10 shows a first try I produced in a few hours, wherein rings of alternating darkness would indicate years (reminding the viewer of annual age rings in tree trunk cross-sections) and the omitted points would indicate the submissions that were awarded prizes.

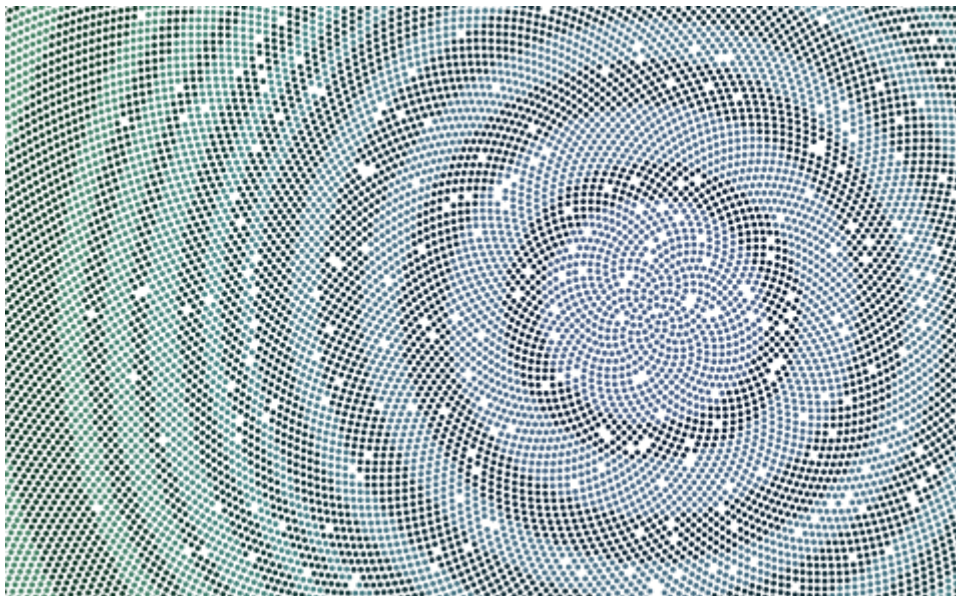* See *http://en.wikipedia.org/wiki/QR_Code*.

Figure 13-10. *Submissions as dots, packed like sunflower seeds*

Despite its visual complexity, the underlying procedure for creating these types of arrangements can be described by simple rules: for placing the *n*th point, choose a radius of the square root of *n*, multiplied by a constant scaling factor. The angle at which the point is placed is the angle of the preceding point, incremented by the golden angle (2*pi/*phi* = ca. 137.5 degrees).

To distribute the points in a homogenous and uniform manner, it is very important to use precisely this number: if we used, say, 137.4 degrees, the characteristic double spiral would be replaced by spirals in only one direction and the point distances would begin to vary. Using the golden angle, we can add points indefinitely, and each point will be a uniform distance from its neighbors. Why is this the case? Each rational number we pick for dissecting the circle will result in a repetition of angles sooner or later. In the simplest case, if we always move a half-circle ahead, we will only end up with two different angles. It can be shown that for any rational fraction, there will be a repetition and thus a limited set of angles used. Accordingly, if we want to optimize the filling and the distribution of points, we have to use an irrational number—ideally, the most irrational number there is (that is, the one that is the least well approximated by a fraction). This number is *phi*, the number representing the golden section.

The Final Product

Having found a guiding visual principle, many of the open questions and possible combinations were now naturally reduced to what worked within the self-imposed constraints—for instance, the principle dictates circular shapes for all groups of items. As the category distributions were of importance in all the perspectives we discussed, we decided to color-code the categories across all the visualizations to be displayed, with identical colors used for groups of categories that could reasonably be treated as "families" (for instance, the categories in the field of computer animation and film are all shown as orange). In addition, I introduced a shape encoding to indicate whether a submission had received a prize or not (circles for nonwinners, diamonds for winners).

As discussed earlier, on a conceptual level, I became interested in the relation of the totals and sums to the individual submissions. Consequently, I looked for a way to incorporate this information into the final visualization. After some unsuccessful experiments with putting additional labels for the total counts around the circles and overlaying the count numbers on top of the circles, which led to quite cluttered displays, I found a much more satisfying alternative: the numbers could actually be created by the dot pattern itself! As the decision to color-code the categories ruled out all modifications to the points themselves, I decided to skip all positions in the sequence that would occlude the number, if it were overlaid on the circle (see Figure 13-11). That dot would simply go in the next available precalculated position, so the total number of dots would remain the same but the circle size would increase marginally. Obviously, this principle only works for circles with enough dots to create the number; accordingly, the number is only displayed for circles containing a minimum of 100 items.
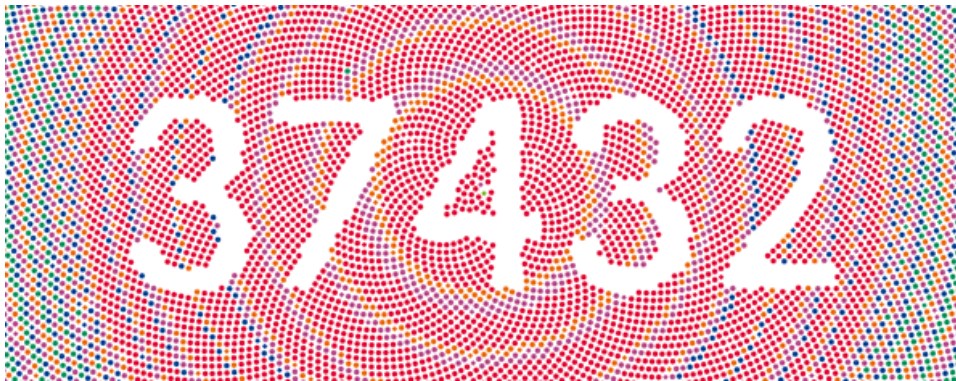

Figure 13-11. *Numbers created by skipping points in the placement sequence*

All Submissions

Figure 13-12 shows all submissions to the Prix Ars Electronica over the last 22 years. Resembling a tree trunk cross-section, the oldest submissions are located at the center, surrounded by the more recent ones. This constitutes the starting point for all the other graphics, each of which is a split-up version of this one, with data analyzed according to different criteria.
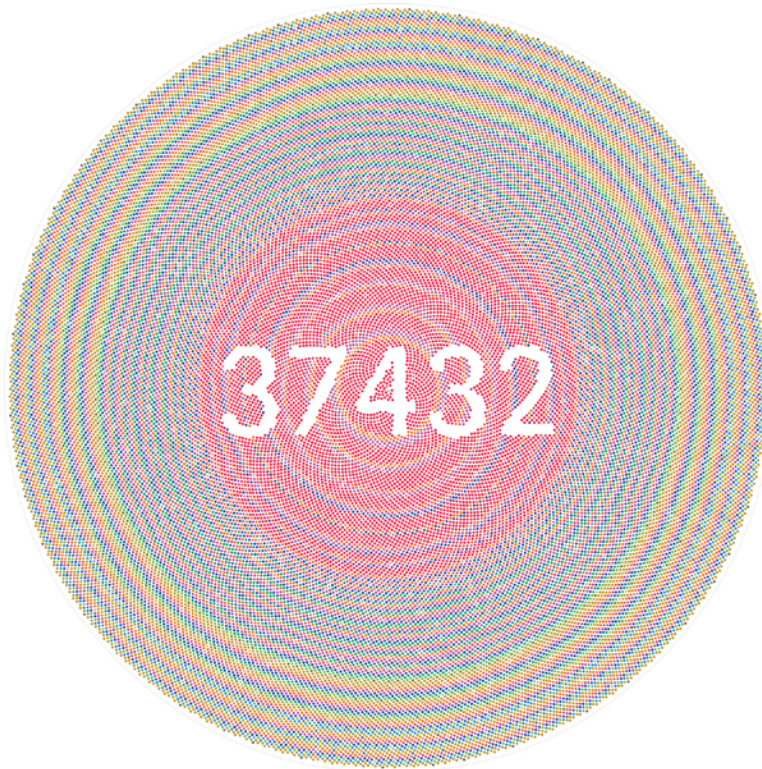

Figure 13-12. *All 37,432 submissions, colored by category and arranged from inside (least recent) to outside (most recent) by submission year*

By Prize

The diagram shown in Figure 13-13 is enough to motivate the whole project: splitting the submissions by the prize received (or not) reveals that only 4% of all submissions have received an honorary mention, a distinction, or a Golden Nica. The remaining 96% of submissions remained invisible—up to now. For this, and all the following, more analytical views, I decided to show the category distribution within the groups of data in a pie chart fashion, in order to avoid the perceptual distortions introduced by the concentric rings in the overview graphic.
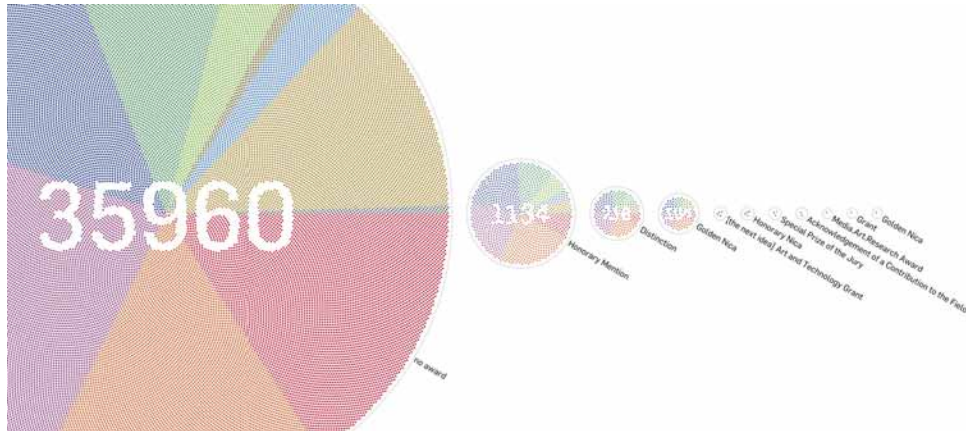
Figure 13-13. *Submissions by prize*

By Category

Figure 13-14 shows a quantitative analysis of the submissions by category. At the same time, it provides a sense of the fraction of awarded projects per category in a fainter section of the pie, composed of diamond shapes in the right part of each circle. It shows, for instance, that the computer graphics category has the highest number of submissions (per single category), but that a low number of prizes has been awarded per submission (a result of the fact that the category has been around for only seven years). Following Wang et al. (2006), the layout of the circles was calculated using Flare's CirclePackingLayout algorithm.
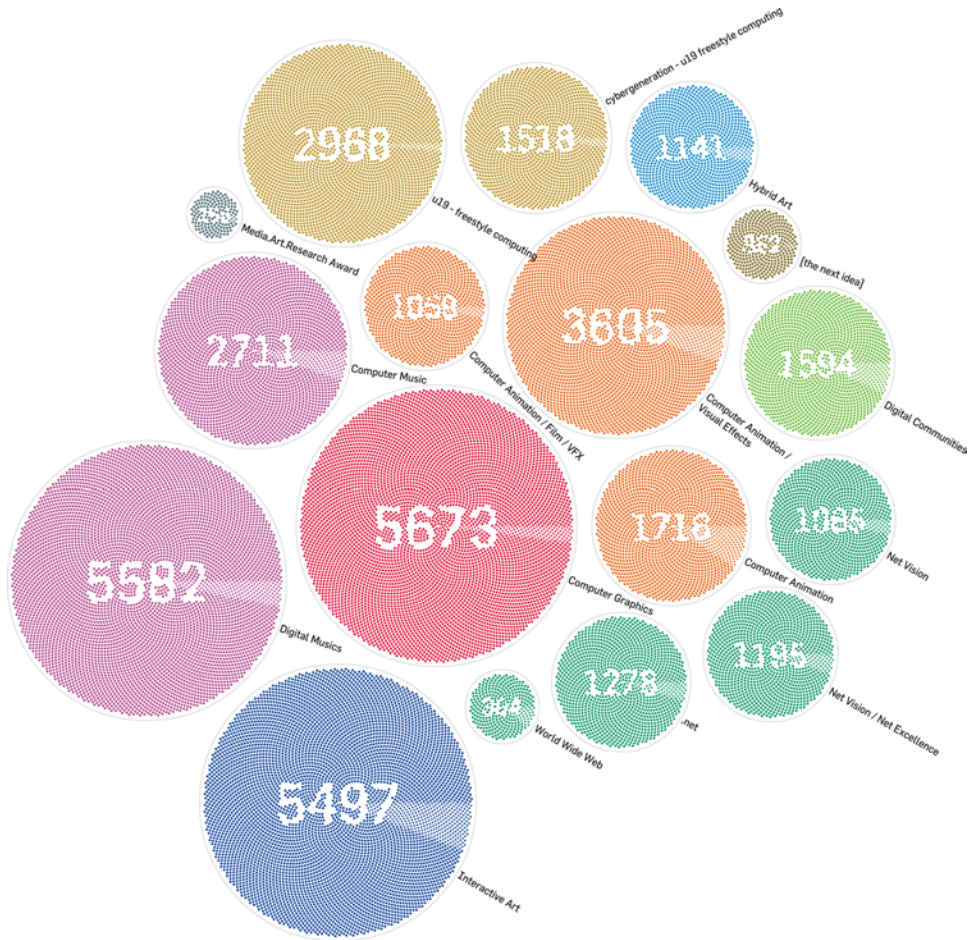
Figure 13-14. *Submissions by category*

By Country

Figure 13-15 shows a map of the submitters' countries of origin. Inspired by the *New York Times*'s map of Olympic medals,* the layout is calculated with a physical rigid body model and attempts to approximate the exact locations, while avoiding circle overlap (see Figure 13-16 for snapshots of the iterative optimization process).

* See *http://www.nytimes.com/interactive/2008/08/04/sports/olympics/20080804_MEDALCOUNT_MAP.html*.
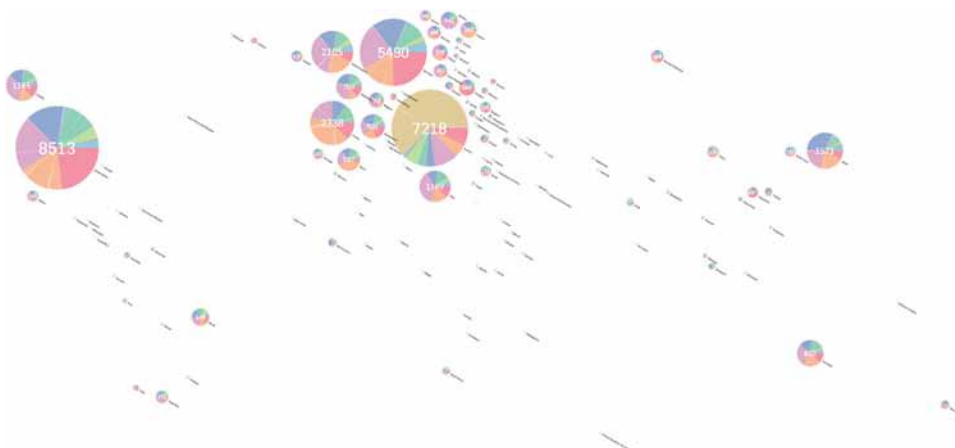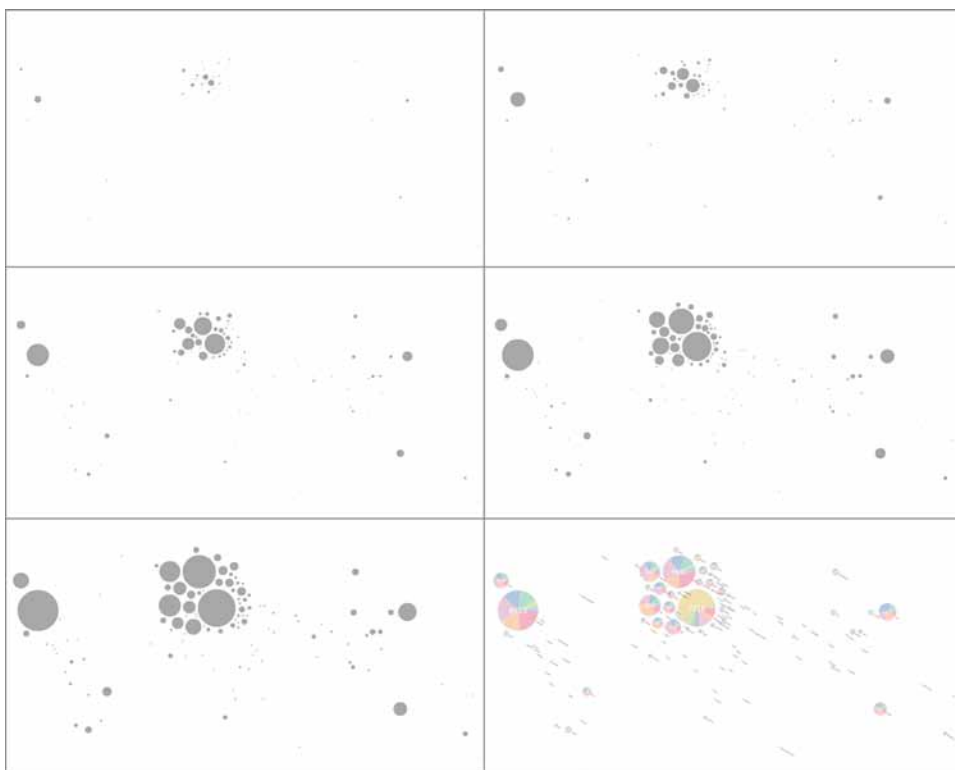
Figure 13-15. *Submissions by country*

Figure 13-16. *Snapshots of the iterative map optimization*

To get coordinates for the country names I used the online application *mapspread,*[*] which allows users to batch-query tabular data for geocoordinates. However, some manual correction was required, as some of the country names could not be resolved (the Eastern European political landscape, in particular, has changed quite a bit over the last few decades), and others were ambiguous: in fact, even in the final version, the label "Georgia" was mistakenly placed next to the United States instead of over the Eastern European country located between Russia and Turkey.

Inspecting the map in detail reveals the European/U.S.-centric nature of media art, with very few contributions from South America, Africa, Russia, or Asia (with the exception of Japan). Historically, a large number of submissions from France and Spain have been made in the field of computer animation and film (orange). Italy, Sweden, and the UK show a tendency toward music categories (purple), while Japan seems more into interactive art (blue). In contrast, Germany and the U.S. leaned toward computer graphics (red), at least in the early years of the festival. Almost two-thirds of Austria's submissions have been in the (Austrian-only) U19 categories.

By Year

The sequence of pie charts in Figure 13-17 shows a clear division of the prize history in three eras. In 1995 there was a sudden decrease in submissions, coinciding with the discontinuation of the computer graphics category and the introduction of the World Wide Web category. One possible explanation for this drop is that it was more common to submit multiple pieces per year in the computer graphics category. The years after 2004 show a stronger diversification in categories and a sudden increase in submissions, largely due to the introduction of the U19 categories for Austrian artists under 19 years of age.

Figure 13-17. *Submissions by year*

By Year and Category

Figure 13-18 shows a matrix version of the timeline, to allow inspection of the development of individual groups of categories. In both color-coding and row selection, we decided to group corresponding categories even if their titles were changed over the years. (Conversely, it should be noted that some categories whose names did not change had different orientations in different years.) Compared to the single-year chart, this version makes it easier to see how animation/film, music, and later interactive art have become the long-term backbones of the Prix.

* See *http://mapspread.com*.

Figure 13-18. *Submissions by category and year*

Exhibition

"Mapping the Archive" was located at the history lounge exhibition in the
Brucknerhaus, and featured six different data visualization perspectives created by
Dietmar Offenhuber, Evelyn Münster, Jaume Nualart, Gerhard Dirmoser, and me
(Figure 13-19).[*]

Figure 13-19. *The poster in the exhibition*

To facilitate the discovery of individual stories in the data, we added little annotation
arrows to highlight interesting facts (Figure 13-20). Visitors were also encouraged to
add their own annotations, resulting in a couple of interesting questions and remarks.

---

[*] All visualizations are documented online at *http://vis.mediaartresearch.at*.

Figure 13-20. *Sticky arrow notes with handwritten annotations*

Conclusion

The visualizations presented here were developed over the summer of 2009, through a continuous exchange of ideas and information not only with the technical staff in charge of the archive databases, but also with media art experts commenting on the semantic perspective of the represented information.

I see the work as part of a young tradition of *information aesthetics*.* The scientific discipline of information visualization is usually concerned with characterizing methods of visual mapping in general, and optimizing the readability and understandability of the resulting visualizations. Information aesthetics builds on results from that area; however, being a design discipline, it strives to find a sensory representation of information based on a specific dataset that not only is useable and readable on the explicit data-representation level but, in addition, increases the "propositional density"† of the design piece—in short, the evocative character of a visualization and what can be read "between the lines." This approach places the discipline between the traditional fields of information visualization, user interface design, and art.

I hope this chapter demonstrates some key features of the discipline. First, it is important to look at the process of creating information aesthetic works. In my experience, it is crucial to work with realistic data, already even in the early stages of the design. In principle, many visualization ideas developed by theory early on could work well on

---

* A term first coined by Lev Manovich and explicated in detail in Lau and Vande Moere (2007).
† As defined by William Lidwell (2009).

real-life data structures, but whether they deliver interesting information and are useful for answering the questions in mind—or for provoking new questions—can only be determined when working with the actual data. Developing visualizations has to be a bootstrapping process: you must use them early on to understand which visualizations and data treatments to pursue further. In our case, early visualization experiments with standard tools put us in a position to understand which data fields to use and which combinations of data "smelled" interesting, and provided a good basis for discussing the design features of the future visualization with reference to concrete, realistic examples. If the designer does not allow his own visual explorations to change his mind on the way to the final product, chances are high that the result will only state the obvious, without provoking new questions or revealing interesting stories.

Moreover, it is crucial to be aware of the semantic context of the information displayed, and the semiotic character of the final piece. To give an analogy, in linguistics, the field of semantics is concerned with the study of sentence meaning as it can be constructed from its constituents and their combination. However, it is widely acknowledged that language can only be fully understood by also looking at pragmatics: the study of how language is actually used in a social context. What is the connotation of a word or expression? What associations does it evoke? And what is expressed by not saying anything? What form of expression is expected in a given context, and what goes against the norms?

Much effort has been invested in understanding the syntax and semantics of visual language for information presentation, and now information aesthetics is opening the door for an investigation of the pragmatics of visual language. In the work presented here, for instance, the chosen visual principle was born out of the inherent tension induced by approaching a complex social phenomenon from a purely quantitative angle. What statement were we making in breaking down a tremendously rich and varied dataset, representing 22 years of media art history, in all its facets, into "a couple of numbers"? The form of the visualization tries to capture this tension and resolve parts of it.

Given these considerations, the notion of "aesthetics" in visualization is about much more than "pretty pictures." Surely, joy of use is an important and long-underestimated factor—on many occasions, research on the user experience has shown the importance of interacting in pleasant, stimulating environments. But, as Steve Jobs famously remarked, "Design is not how it looks and feels, but how it works." A truly aesthetic visualization, in addition to being beautiful, "works" by expressing inexplicit features of the phenomenon at hand, and inviting the user/viewer to explore a rich and multifaceted world.

As a final remark, looking at the meaning and context of the information presented in a visualization, one point is often neglected (even in the work presented here): how can we characterize that information in the larger scheme of things? Could we find explanations for some of the patterns observed by connecting to external databases? In the Ars Electronica example, it might have been informative to compare, for instance, submission statistics per country with more information about each country. Is the number of submissions correlated to economic power? Or digital literacy? Or other, less obvious factors? As more and more open data sources for these types of information become available, it is increasingly important to provide the proper context and baselines for actually understanding the significance of patterns arising in the datasets we analyze and present.

Acknowledgments

References

Keim, D.A. 2000. "Designing pixel-oriented visualization techniques: Theory and applications." *IEEE Transactions on Visualization and Computer Graphics* 6, no. 1: 59–78.

Lau, Andrea, and Andrew Vande Moere (2007). "Towards a model of information aesthetic visualization." In *Proceedings of the International Conference on Information Visualisation*. Washington, DC: IEEE Computer Society.

Lidwell, William. 2009. "More with less." *ACM interactions* 16, no. 6: 72–75.

Wang, Weixin, Hui Wang, Guozhong Dai, and Hongan Wang. 2006. "Visualization of large hierarchical data by circle packing." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM Press.

Wattenberg, Martin. 2005. "A note on space-filling visualizations and space-filling curves." In *Proceedings of the 2005 IEEE Symposium on Information Visualization*. Washington, DC: IEEE Computer Society.