

# ID: An Illustrator Dataset - Data Collection Details

Compiled by Jaleen Grove, Ph.D, [Jaleen@gmail.com](mailto:Jaleen@gmail.com) <http://jaleengrove.com/index.htm>

I gratefully acknowledge the support of the DB Dowd Modern Graphic History Library on this project.

Data is compiled and visualized at

[https://public.tableau.com/profile/jaleen3968#!/vizhome/IDTheIllustratorDataset/1\\_Introduction?publsh=yes](https://public.tableau.com/profile/jaleen3968#!/vizhome/IDTheIllustratorDataset/1_Introduction?publsh=yes) If the link stops functioning, visit my website or email me for an update.

The data visualizations and bibliography took a very long time to make. It is freely available for use but I'd appreciate an acknowledgment in anything you blog, present or publish that was helped by it.

Thanks!

June 21, 2018

In any data project, decisions have to be made about how and what to collect. Users of *ID* need to know how much they can trust it, and how well it matches the choices governing other datasets elsewhere that they may want to compare. In data collection it is important to set rules and stick to them for consistency, although inevitably exceptions to the rule have to be accommodated. The following sections document the decisions, rules, and exceptions that characterize *ID*.

## Data Literacy

It's easy to get sucked in to believing what you see. Data visualizations can make all kinds of implications through mere use of colour, the slant of a line, proximity of marks. Research how to read visualizations critically.

The "truth" of what you see in *ID* is dependent on how valid my selection of spreadsheet entries is and how few errors I or the original authors made in typing. That it is a relatively large corpus balances out mistakes and flukes, but be especially suspicious about findings based on the smaller figures, such as for the subdiscipline Music.

The impression of illustration history that this data gives is accurate according to the publications the data came from, but should not be taken as THE history or canon. A completely different dataset might yield a contrasting set of "greats". The value of *ID* is not that it has the last word, but that it can act as a yardstick with which other datasets can be compared—not to find out whether the second dataset "measures up", but rather so that differences can be identified, which in turn can illuminate the values, deviances, and biases of both. Solid research needs to account for such things and this is a tool to help do it.

## Overview

### *Guiding purpose*

This project began as an inquiry into how archives construct knowledge, and an investigation into women's inclusion or exclusion in the profession of illustration. Gender was therefore recorded for writers and illustrators, and many of the resulting visualizations compare male and female representation. A paper that critically examines the findings of this data is being prepared.

### *Basic elements*

- The underlying Excel spreadsheet has 19,023 entries.
- Publications amount to 121.
- Editors, single-authors, and collaborating authors of entire publications amount to 208; the number of authors writing in parts of edited volumes has not been tabulated.
- The breakdown of writers by sex is 159 males, 33 females, and 16 unidentified.
- The number of illustrators' names appearing in this corpus is 7225, but this is not quite the number of actual illustrators because in some instances it could not be determined for sure whether similar names were for the same individual, as in JA Adams and John A. Adams. Also, as explained below, "illustrators" includes a few non- and semi-illustrators.
- The breakdown of illustrators by sex is 5914 males (81.85%), 1,791 females (10.95%), and 520 unidentified (7.20%).
- Columns record publication title, chapter or part's title, subdiscipline, overseeing editor(s), writer(s), illustrator, and the sexes of the writers and illustrators where known: 'U' designates Unknown/unidentified or a collaboration of man and woman.

### *Subdisciplines*

- Subdisciplines are *Undifferentiated, Non-US/UK, Children's, Advertising, Science, Avant-garde, Cartoon/Satire, Western, News, Music*
- *Avant-garde* designates texts prioritizing Arts and Crafts "decorative" and "book beautiful" efforts, high modernism, *livres d'artiste* and *éditions de luxe*. What holds these stylistically differing categories together is a shared desire to make books different from the common run, privileging the gestalt of the whole and giving more power and prestige to the illustrator/artist.
- Subdiscipline is not assigned based on my own knowledge of the illustrators. Subdiscipline is assigned if 75% or more of a whole publication or whole chapter seems devoted to a specialization. In deciding on subdiscipline classification, if it isn't obvious from the title and contents (different writers spoke of 1890s Arts and Crafts illustrators as either mainstream illustration or avant-garde, for instance), the writer's rhetoric is consulted.

- Subdiscipline is an imperfect way to filter since so many texts combine types of illustrators or refer to out-of-subdiscipline illustrators for comparison, and since many illustrators cross subdisciplines. The subdisciplines *Non-US/UK*, *Avant-garde* and *Children's*, however, are fairly clear-cut. The ability to see certain illustrators cited in the context of a given subdiscipline, whether they belong to it or not, is informative.
- *Undifferentiated* and *Non-US/UK* are not subdisciplines but rather catch-all categories of all the subdisciplines, with an emphasis on mainstream trade book and magazine illustration (fiction and reportage). The former includes non-English and non-American illustrators who were mentioned in chapters that give general overviews. The *Non-US/UK* category is for entire chapters devoted to other countries.

### *Dates*

- To make the visualizations work, all texts have to be given a unique date in the exact same format; duplicating dates must be avoided (not possible if two publications were issued the exact same day).
- All texts are assigned a year, month, and day. Dailies and weeklies are given their precise date. Monthlies are assigned the first day of the month issued; or if a duplicate, then the 7<sup>th</sup>, 15<sup>th</sup>, or 21<sup>st</sup> as needed. Books are given publication dates of January 1 as a rule; books sharing the same year have been arbitrarily assigned the first of another month.

### General guidelines used in data entry: inclusions

- The ideal is to include every known English-language book, article, exhibition catalogue, and syllabus that surveys multiple wood engravers, illustrators and cartoonists past and present, that was accessible to English-speaking readers. This is as close as could be achieved in the time available.
- Each text had to be published between Jan. 1, 1830 and Dec. 31, 1970. Perhaps one day the range will be increased to 1990; for now, time constraints forced this date-limit.
- This corpus has been located by going through the bookshelves of the Washington University Special Collections, which has a strength in book arts, and by locating missing texts identified in the bibliographies of those books; and by querying databases such as American Periodicals Online, Hathi Trust, Arts Search, and Google News. The one syllabus by Harve Stein resides in the archives of the Society of Illustrators and at RISD.
- Surveys, exhibition catalogues, and critical reviews of illustrators are prioritized, where the writer has exercised judgment over a category about who to include. Books on *aspects of illustration* and on *illustrated books* as opposed to on *illustrators* are included if illustrators are named and the scope seems to merit inclusion. Manuals for

practitioners and works of criticism are included if they give sufficient scope of illustrators held up as exemplars, such as Knaufft's 1899 *Drawing for Printers*.

- Only first editions are consulted, in order to pinpoint when a name was first introduced to the history discourse. An exception is made for the three *Illustrators of Children's Books* volumes, which each introduce substantial numbers of new names by new writers; and for Sinclair Hamilton's *Early American Book Illustrators . . .*, Volume II, where only the new names have been recorded since it is simply an extension to Volume I. This means that new names added to revised second editions have not been recorded.
- Each book chapter is counted as an individual text (making several texts per book); likewise, articles belonging to a series spread over several magazine issues (Part I, Part II, etc.) also count as multiple individual texts.
- When chapters were very brief they were combined and treated as if they were one chapter.
- Chapters in highly specialized, single-author books that largely repeat the same names over and over are treated as if they are one chapter (*White's English Illustration: The Sixties*; *Strachan's The Artist and the Book in France*, e.g.).
- Only the first mention and the first artwork shown for an illustrator in each chapter is recorded.
- "Illustrator" here includes people normally understood to be Old Masters, engravers, cartoonists, graphic designers. Often it is impossible to differentiate the definition of illustration from art, printmaking, cartooning, designing, and people wore multiple hats. It was important to throw the net wide so as not to second-guess who the writer considered the illustrators to be, since the project is primarily to meant to capture writers' biases. Also, I anticipate value will be gained from seeing what non-illustrators were being cited in the context of illustration.
- Ruskin, Thackeray, and others who both wrote and illustrated are only recorded when the author refers specifically to their artwork. An exception is William Morris, since usually it could not be determined just what aspect of Morris was meant; and because I thought it important to see how often he was mentioned no matter what.

#### General guidelines used in data entry: exclusions

- Indexes and directories that attempt to catalogue every single illustrator, with no discrimination of quality. An exception was unintentionally made for Sinclair Hamilton's *Early American Book Illustrators and Wood Engravers 1670-1870*.
- Articles that only feature 'house' illustrators (ones regularly employed by the publisher of the consulted text) have been eliminated as too biased, but ones by third parties are kept (*Halsey's Illustrating for The Saturday Evening Post*).

- Articles and monographs on single illustrators are left out.
- Chapters pertaining entirely to illustrators prior to 1800 are omitted, since the focus is on Modern illustration, not Early Modern. Chapters that discuss earlier illustration with nineteenth and later centuries are recorded in full, however.
- Entire publications concerning only illustrators outside of Europe and United States are excluded, but publications surveying the US and Europe alongside the rest of the world are kept.
- Chapters focusing on typographers or designers are omitted, but such practitioners are recorded when mentioned in illustration-focus chapters, if that person is shown to be designing typographic ornaments and layouts that go beyond plain typography (ie, species of decorative illustration).
- Printers who are not said to engrave are excluded; those such as Evans who also engraved are kept
- Pure landscape artists have been omitted when their pictures are merely reproductions of paintings, for all that they may 'illustrate' a locale, but this was followed irregularly since it wasn't always clear that the person only did landscapes; and often the writer did consider the landscapes illustrations (in which case they were kept).
- New names added to revised second editions have not been recorded.
- Unattributed illustrations are excluded from the "Art Printed" tabulation.

### Names, spelling, and data cleaning

- Software cannot correctly tabulate data if there are differences in spelling or format, such as full names versus initials, or varying spaces around hyphens, so measures were taken to enforce standardization.
- Name spellings follow the most-used spelling in the corpus or the current Wikipedia entry.
- There is much variation in the literature and it wasn't possible to tell when people were normally referred to by a short prefix. All have therefore been treated the same regardless of custom and what I believe to be true: *de*, *d'*, *le*, or *von*, etc., are given capitalization and made to begin the last name even if not always used that way historically (e.g., Von Menzel instead of Menzel; Von Schmidt, Harold instead of Schmidt, Harold von).
- Similarly, what a middle name versus a two-part surname is has also been maddening. The rule adopted is to record last one first, followed by first name and next name (e.g., Houghton, Arthur Boyd, not Boyd Houghton, Arthur).
- Women's married names take precedence, with the maiden name following the given name (Biller, Olive Allen), even if professional work was done under the maiden name. Tableau's search boxes return all instances of the spelling entered, so this will not

prevent searching. In case some married versus maiden names were not caught, it would be wise to search for both separately and see if double entries show up (please notify me!).

- Pen-names, nicknames, and titles usually appear in brackets after the proper name (e.g., Geisel, Theodore (Dr. Seuss)); although in some cases where the pen-name is indisputably better-known, it has been kept as the primary name (e.g., Gavarni (Guillaume Sulpice Chevalier)).
- In some cases people were referred to by the title M. for Monsieur, but it was not always clear whether it was instead a first initial, so some individuals may inadvertently appear as Surname, M.
- To ensure exact standardization, data were cleaned with the freeware OpenRefine to locate and replace similar monikers, which caught typos and variations. Then Excel columns were sorted alphabetically by name and inspected manually, which caught further variations.
- It is possible that some names that ought to have stayed different were merged, but the advantage of catching all possible mentions outweighs the disadvantage of losing names that only appeared once or twice (more singular name-spelling being more likely to be misspellings than actual separate individuals). Where there was doubt, research was conducted to determine whether to merge or not. Often, the true identity of the illustrator had to be determined by looking at the context of the chapter they appeared in and the other people mentioned there, and comparing to whether a similar name typically appeared in that same context in other books.
- Sex of the illustrator was mostly determined by name. Illustrators known only by initials were googled. Tom Seidmann Freud (who was born Martha Freud) is an example of the pitfalls of guessing sex based on name.
- Sex was also determined by whether the author used male/female pronouns. But in old usage, all illustrators may be referred to as “he” even when the identity is unknown. If it was quite clear from the writer’s discussion that they were making an assumption, I marked the sex U (Unidentified).
- Special mention must be made of the couple Lewitt-Him (Jan Lewitt and George Him), who are represented as collaborators separately from their individual mentions. Those wishing to inspect all the mentions for either man should add the joint-mentions to the individual’s. In most other cases, which amounted to few, the collaborators are given separate entries, especially if of different sex (e.g. Leo Dillon and Diane Dillon). Some collaborators, who do not appear individually too, remain combined if their sex is the same.