



Disseminate your Research: Style Does Matter

Pierre Senellart





Substance vs Style

- The **content** of your research is the most important thing...



Substance vs Style

- The **content** of your research is the most important thing...
- ... but the way you **present** it can make the difference.



Substance vs Style

- The **content** of your research is the most important thing...
- ... but the way you **present** it can make the difference.



*Accept or Reject?
Should I read this paper?
Is this worth funding?*



The role of writing, presentation, and dissemination

- **Serving** the substance of your research
- **First** and **semi-conscious lasting** impression on an article, a presentation a work
- **Facilitating** reading, **making** your proofs **clear**, **highlighting** your results, putting **into context**, providing **supporting** material
- In science, style should **never** go **before** substance
- All of this is only my personal view 😊



What I will not talk about

English grammar. Very important, but I am neither a native speaker nor an English teacher.

Writing style. Important as well, but I am not a literature major.

Oral presentation skills. Practice this!

Content, domain-specific. Ask your advisor!



What I will not talk about

English grammar. Very important, but I am neither a native speaker nor an English teacher.

Writing style. Important as well, but I am not a literature major.

Oral presentation skills. Practice this!

Content, domain-specific. Ask your advisor!

Examples come from PIKM 2012 articles (except for charts, from CIKM 2013 articles).



Style Does Matter

Typography of Research Articles

Don'ts

Do's

Visual Display of Quantitative Information

Online Presence

To Go Further

BADOD
TYPOGR
GRAPHY
IS IS
EVERYSI
BL**WHERE**

**GOOD TYPOGRAPHY IS INVISIBLE /
BAD TYPOGRAPHY IS EVERYWHERE**

PROVIDED A SERIES OF ALPHABETS FOR THE DESIGNER TO CHOOSE FROM
BASED ON THE IMPORTANCE OF VISUALS AND DESIGN. THE DESIGNER
AND VISUALS. VISUALS ARE IN THE END AND THE END OF THE END
IS THE END. THE END IS THE END. THE END IS THE END. THE END IS THE END.
AT THE END OF THE END IS THE END.



Craig Ward



Why typography conventions matters

Typography is the art and technique of arranging type in order to make language visible.

Wikipedia



Why typography conventions matters

Typography is the art and technique of arranging type in order to make language visible. *Wikipedia*

- Typography has developed since the invention of the **printing press**.
- A number of **typographical conventions** have emerged, beyond grammar and language usage conventions.
- Sometimes chosen for **readability**, sometimes **fully arbitrary**.
- Because we are used to read printed documents (or documents typeset on a computer screen), we have assimilated these conventions.
- Any deviation to these conventions is noticed and **distracts** from the actual content.



Should we always respect conventions?



Should we always respect conventions?

- Not necessarily!



Should we always respect conventions?

- Not necessarily!
- Some conventions may actually hurt readability of some kinds of typeset material.



Should we always respect conventions?

- Not necessarily!
- Some conventions may actually hurt readability of some kinds of typeset material.

Example (U.S. usage)

Periods and commas always go inside quotation marks, even inside single quotes.



Should we always respect conventions?

- Not necessarily!
- Some conventions may actually hurt readability of some kinds of typeset material.

Example (U.S. usage)

Periods and commas always go inside quotation marks, even inside single quotes.

The keyword query used in the experiment was “olympics 2012.”

or

The keyword query used in the experiment was “olympics 2012”.



Should we always respect conventions?

- Not necessarily!
- Some conventions may actually hurt readability of some kinds of typeset material.

Example (U.S. usage)

Periods and commas always go inside quotation marks, even inside single quotes.

The keyword query used in the experiment was “olympics 2012.”

or

The keyword query used in the experiment was “olympics 2012”.

- But:
 - Know the conventions
 - Have a good reason to deviate when you do
 - Be consistent

Be consistent!



Outline

Style Does Matter

Typography of Research Articles

Don'ts

Do's

Visual Display of Quantitative Information

Online Presence

To Go Further



What's wrong with this?

”‘dangerous’” or ”‘non-dangerous’”



What's wrong with this?

”‘dangerous’” or ”‘non-dangerous’”

Abuse of typography conventions

- Weird usage of quotation marks: “...”



What's wrong with this?

Several problems remain to be addressed in future research. Additional work is required to improve the integration and exploration of source code repositories and a database's schema (e.g. parallel processing techniques).



What's wrong with this?

Several problems remain to be addressed in future research. Additional work is required to improve the integration and exploration of source code repositories and a database's schema (e.g. parallel processing techniques).

Abuse of typography conventions

- Missing comma after “e.g.”
- Improper double spacing after “e.g.”



What's wrong with this?

For both input topics 'CAD' and 'Kashmir', we observe a very small percentage of users (CAD : 17% , Kashmir : 0.8 %) which have the input topic in their bio. But, the



What's wrong with this?

For both input topics 'CAD' and 'Kashmir', we observe a very small percentage of users (CAD : 17% , Kashmir : 0.8 %) which have the input topic in their bio. But, the

Abuse of typography conventions

- Inconsistent spacing before “%” (the preferred writing is without a space)
- Extraneous space before colons and comma



What's wrong with this?

type conversion, without the need to change the original data. For example it is possible to define Attribute 1 as a sum of Data Field 1 and Data Field 2. It is also possible to use a data field in more than one mapping. Although they can be bound to the same data,



What's wrong with this?

type conversion, without the need to change the original data. For example it is possible to define Attribute 1 as a sum of Data Field 1 and Data Field 2. It is also possible to use a data field in more than one mapping. Although they can be bound to the same data,

Bad line break

- Breaking a line between “Data Field” and “1” hurts readability



What's wrong with this?

for each $j = 1, 2, \dots$, the probability θ_{ji} related to the activity j are learned with a HMM of states x and observations y ,



What's wrong with this?

for each $j = 1, 2, \dots$, the probability θ_{ji} related to the activity j are learned with a HMM of states x and observations y ,

Bad line break

- Breaking a line between “activity” and j hurts readability



What's wrong with this?

$$irf(k) = \log\left(\frac{n}{\hat{n}} + 1\right) * \log\left(\frac{|E|}{\hat{e}} + 1\right)$$



What's wrong with this?

$$irf(k) = \log\left(\frac{n}{\hat{n}} + 1\right) * \log\left(\frac{|E|}{\hat{e}} + 1\right)$$

- * instead of \times (or \cdot) for multiplication
- Improper kerning in the function name, should be: *irf*



What's wrong with this?

$$Dist_{max}(s, l, t) = \max\{Dist(d_i, s, l, t) | d_i \in D_{s,l}^t\}$$



What's wrong with this?

$$Dist_{max}(s, l, t) = max\{Dist(d_i, s, l, t) | d_i \in D_{s,l}^t\}$$

- *max* should be `max`: this is an operator name, not a variable name
- Improper spacing around the `|` operator; it should be typeset using `\mid`, not `|`



What's wrong with this?

Figure 8: The figure shows the scatter plot for betweenness centrality versus degree centrality for the topic-centric community 'Kashmir' ($k = 15$).



What's wrong with this?

Figure 8: The figure shows the scatter plot for betweenness centrality versus degree centrality for the topic-centric community 'Kashmir' ($k = 15$).

- The authors have tried tweaking the size of the caption, probably using a `\small` command, probably to gain space.
- Bad idea to change the default style imposed in the document class, for consistency.
- More importantly, since only part of the paragraph has changed font space, line spacing has not been changed! It is now too large, and no space is saved at all.



What's wrong with this?

```
1: for  $t_i$  in  $T$  do
2:    $\mathcal{C}_{current}^{t_i} = \text{search\_bio}(t_i)$ 
3:    $\mathcal{U}^{t_i}, \mathcal{B}^{t_i}, \mathcal{G}_{current}^{t_i} \leftarrow \text{expand}(\mathcal{C}_{current}^{t_i})$ 
4:   while Continue_Condition( $\mathcal{R}_{max}[\dots]$ ) is TRUE do
```



What's wrong with this?

```
1: for  $t_i$  in  $T$  do
2:    $C_{current}^{t_i} = \text{search\_bio}(t_i)$ 
3:    $\mathcal{U}^{t_i}, \mathcal{B}^{t_i}, \mathcal{G}_{current}^{t_i} \leftarrow \text{expand}(C_{current}^{t_i})$ 
4:   while Continue_Condition( $\mathcal{R}_{max}[\dots]$ ) is TRUE do
```

- Bad kerning inside subscripts. One has the impression that c , u , r , r , e , n , t are individual variables that are multiplied.
- “current”, “max” are not variables, they are labels, they should typeset in upright font, not italics.
- Inconsistency in the use of \leftarrow and $=$ for assignment.



What's wrong with this?

[19] T. N. S. K. P. d. . N. T. Joe Conway, Dirk Eddelbuettel. R Interface to the PostgreSQL Database System, 2012-01-29. <http://cran.r-project.org/web/packages/RPostgreSQL/-index.html>.



What's wrong with this?

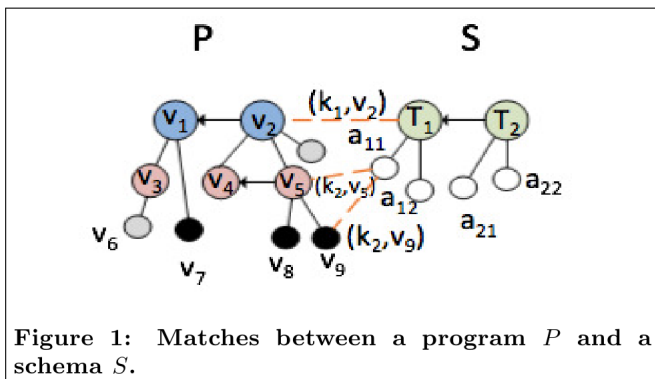
[19] T. N. S. K. P. d. . N. T. Joe Conway, Dirk Eddelbuettel. R Interface to the PostgreSQL Database System, 2012-01-29. <http://cran.r-project.org/web/packages/RPostgreSQL/-index.html>.

Weird formatting of bibliography

- Probable improper use of BibTeX syntax (“and” keyword)
- A URL should not be hyphenated when spanning line breaks

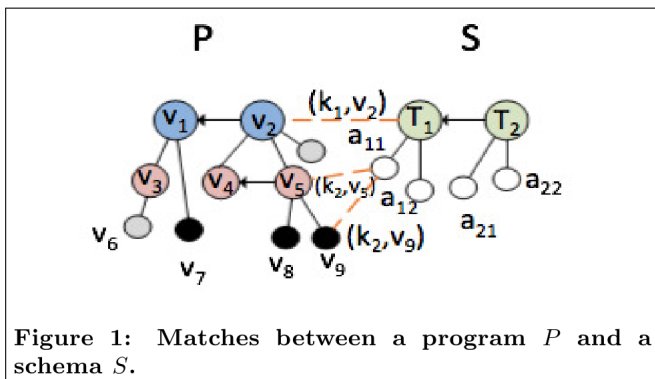


What's wrong with this?





What's wrong with this?

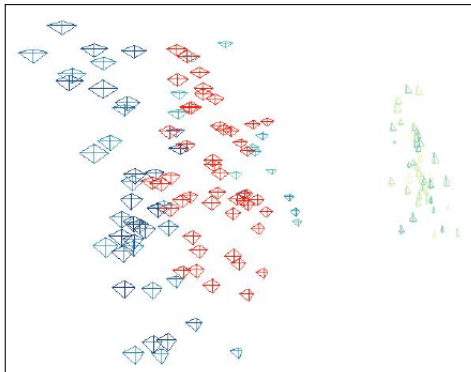


Clashing image

- Low-resolution bitmap image
- Inconsistency between image typeface and text typeface

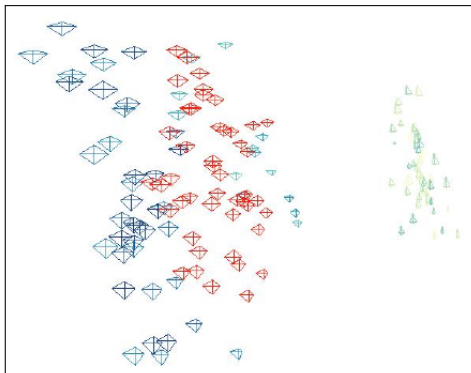


What's wrong with this?





What's wrong with this?



Clashing image

- JPEG compression artifacts



Outline

Style Does Matter

Typography of Research Articles

Don'ts

Do's

Visual Display of Quantitative Information

Online Presence

To Go Further



Punctuation and spacing

French	English
.	followed by a space
,	followed by a double space (usually, when full stop)
;	followed by a space and preceded with a thin unbreakable space
:	followed by a space and preceded with a thin unbreakable space
?	followed by a space and preceded with a thin unbreakable space
!	followed by a space and preceded with a thin unbreakable space



Dashes and quotation marks

\LaTeX	French	American English
-	-	petite-fille, 1979-1981
-	--	grand-daughter
-	--	Je l'ai cru – car il était Perron–Frobenius
-	--	convaincant – et je l'ai 1979–1981
-	---	suiwi. Perron–Frobenius
-	---	— Bonjour
-	---	— Ça va?
“ ”	‘ ’	I believed him—since he was convincing—and I followed him.
“ ”	‘ ’	He said, “Hi”, I reckon.
‘ ’	‘ ’	l’apostrophe
« »	\og \fg	“He said, ‘Hi’, didn’t he?”
« »	\og \fg	Il a dit : « Bonjour ».

In British typography, usage of ‘ ’ and “ ” is reversed.



Line breaks and page breaks

- Pay special attention to line breaks, avoid breaking up **semantic phrases** (a number and its unit, a variable and its type, etc.), so as not to disrupt the flow of reading
- Use nonbreakable spaces (`~`) where needed
 - Given a set `set~X$`
 - As shown in Sections `\ref{a}`, `\ref{b}`, and `\ref{c}`
 - The program ran in `23~ms`
 - As mentioned in `\cite{foobar}`
- Pay special attention to page/column breaks, **avoid a break after a section heading, avoid widows or orphans...** \LaTeX usually good at this for standard classes, but may not be the case with publisher-provided classes
- Use explicit breaks (`\pagebreak`) as needed, as a final step



Stick to a stylesheet

- If using Word or a similar WYSIWYG tool, **rely on styles**, not on explicit formatting instructions (but you should use \LaTeX instead ☺)
- In \LaTeX , **avoid explicit formatting commands** (`\textbf`, `\small`, `\vspace`, `\`, `\noindent...`); they are occasionally useful but tend to be abused
- Stick to **semantic commands and environments** provided by the document class (theorems, sections, paragraph headings, etc.)
- If using a publisher-provided document class, **read the documentation** and carefully follow the instructions
- If no class imposed, **rely on modern document classes** (`scrartcl`, `memoir`, `scrbook...`) rather than on legacy \LaTeX classes (`article`, `report`)



Write semantic math

- Define **macros** for every multiple-letter variable (`\mathit`), and every label (`\mathrm`), operator (`\DeclareMathOperator`) and special notation; this will ensure proper formatting and consistent usage
- Use `\text` for incorporating **text fragments** inside equations
- Use **amsmath** extensions to break equations on multiple lines, to format matrices, modulus, etc. Read the documentation of this package!
- The way the equation is written should (as much as possible) reflect its **semantics**, not the way it is displayed
- Use **inline** (`$ $` or `\(\)`) and **displayed** (`\[\]`) modes as required
- Do **not** use legacy `$$ $$` and `eqnarray`, poor formatting



Harmony between figures and text

- Figures should **not stand out**, should not appear as foreign objects inside the document
- Use **vector graphics** rather than bitmaps
- If you have to use bitmaps, go for **high-resolution, lossless** compression (PNG rather than JPG)
- Use the **same font** as in the text, **same font size, same notation**, same line width as in tables, etc.
- Consider producing **L^AT_EX code for figures**, using tools such as **TikZ** (read the documentation!), metapost, asymptote, Inkscape PDF+L^AT_EX export, dot2tex, etc. This ensures common look, allows the use of user-defined macros in drawing, etc.



Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Don'ts

Do's

Online Presence

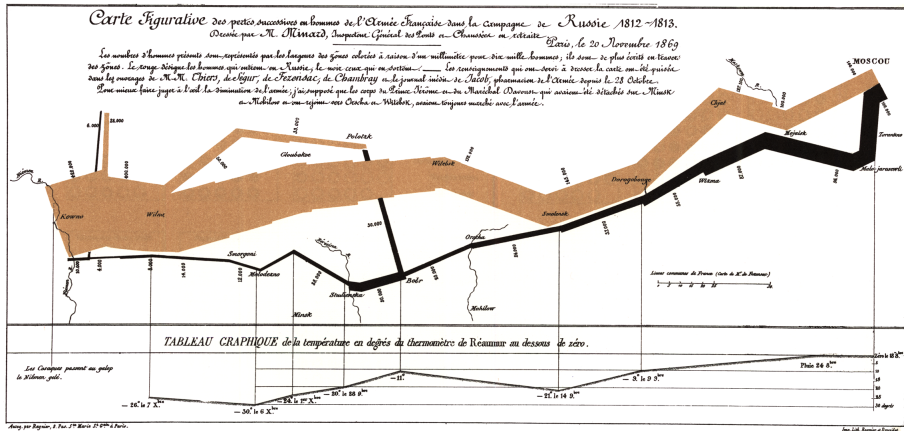
To Go Further

We shouldn't abbreviate the truth but rather get a new method of presentation.

Edward Tufte

We shouldn't abbreviate the truth but rather get a new method of presentation.

Edward Tufte





Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Don'ts

Do's

Online Presence

To Go Further



What's wrong with this?

action	LO vs LE		LO vs CA		LE vs CA	
	χ^2	<i>p</i> -value	χ^2	<i>p</i> -value	χ^2	<i>p</i> -value
BE	9	0.011	9	0.029	18	0.006
FV	9	0.011	9	0.011	18	0.001
LE	13	0.043	6	0.301	27	0.079
QL	36.563	0.064	45	0.039	45	0.039



What's wrong with this?

action	LO vs LE		LO vs CA		LE vs CA	
	χ^2	<i>p</i> -value	χ^2	<i>p</i> -value	χ^2	<i>p</i> -value
BE	9	0.011	9	0.029	18	0.006
FV	9	0.011	9	0.011	18	0.001
LE	13	0.043	6	0.301	27	0.079
QL	36.563	0.064	45	0.039	45	0.039

Badly formatted numerical data

- Numbers not aligned on their least-significant digit
- Inconsistent (and uselessly high) number precision



What's wrong with this?

element	label
Headline	UES
passage within an article	AA
List	LI
Introduction	EL
Picture	BI
Info Box	IB
Charts, tables etc.	IG
Navigation within Wikipedia	IW
Other navigation	BA



What's wrong with this?

element	label
Headline	UES
passage within an article	AA
List	LI
Introduction	EL
Picture	BI
Info Box	IB
Charts, tables etc.	IG
Navigation within Wikipedia	IW
Other navigation	BA

Badly formatted table

- Inconsistent use of lower and upper case
- Not enough contrast between header and table content
- Useless space in between columns, hurting readability
- Missing comma before “etc.”



What's wrong with this?

Musicians	Films	Actors
<i>Johnny Cash</i>	<i>Braveheart</i>	<i>Al Pacino</i>
Bob Dylan	We Were Soldiers	Robert De Niro
Elvis Presley	The Passion of the Christ	Robert Duvall
June Carter Cash	Patton	Marlon Brando
Willie Nelson	Saving Private Ryan	Kevin Spacey
Waylon Jennings	Flags of Our Fathers	Ellen Barkin



What's wrong with this?

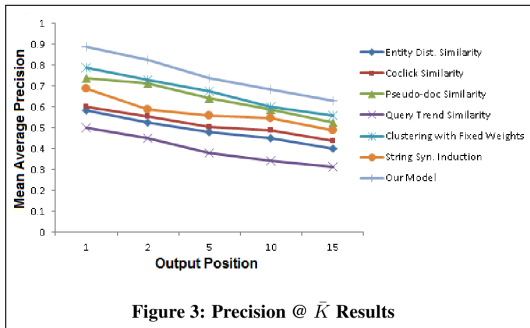
Musicians	Films	Actors
<i>Johnny Cash</i>	<i>Braveheart</i>	<i>Al Pacino</i>
Bob Dylan	We Were Soldiers	Robert De Niro
Elvis Presley	The Passion of the Christ	Robert Duvall
June Carter Cash	Patton	Marlon Brando
Willie Nelson	Saving Private Ryan	Kevin Spacey
Waylon Jennings	Flags of Our Fathers	Ellen Barkin

Hard-to-read table

- Checkerboard effect: vertical rules and repeated horizontal rules hamper readability (the brain sees the lines more than the content)

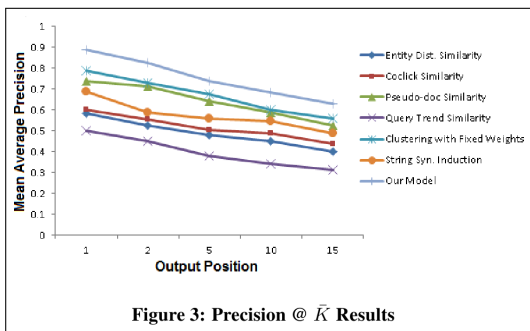


What's wrong with this?





What's wrong with this?

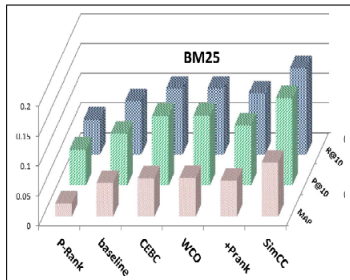


Misleading chart

- Misleading x -axis: trying to follow the evolution of the line chart is meaningless
- Mismatch between chart style and rest of the paper

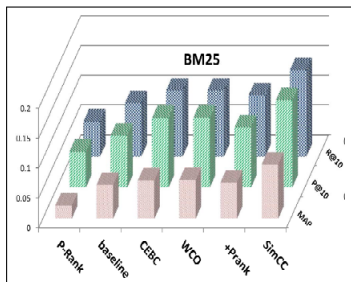


What's wrong with this?





What's wrong with this?



Hard-to-read chart

- 3D plots are notoriously hard to read
- How to compare bars at different depth?
- Pattern fills are distracting to the eye



Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Don'ts

Do's

Online Presence

To Go Further



Beautiful tables: booktabs

- Forget vertical rules: hurt readability
- Forget `\hline`: inflexible spacing
- Use the `booktabs` package for tables

```
\begin{tabular}{cc}  
\toprule  
T&Y\  
\midrule  
a&b\  
b&c\  
b&a\  
\bottomrule  
\end{tabular}
```

T	Y
a	b
b	c
b	a



Beautiful tables: pgfplotstable (1/2)

In a toto.dat file:

```
# Convergence results
# fictional source, generated 2008
level dof error1 error2 info grad(log(dof),log(error2)) quot(error1)
1 4 2.50000000e-01 7.57858283e-01 48 0 0
2 16 6.25000000e-02 5.00000000e-01 25 -3.00000000e-01 4
3 64 1.56250000e-02 2.87174589e-01 41 -3.99999999e-01 4
4 256 3.90625000e-03 1.43587294e-01 8 -5.00000003e-01 4
5 1024 9.76562500e-04 4.41941738e-02 22 -8.49999999e-01 4
6 4096 2.44140625e-04 1.69802322e-02 46 -6.90000001e-01 4
7 16384 6.10351562e-05 8.20091159e-03 40 -5.24999999e-01 4
8 65536 1.52587891e-05 3.90625000e-03 48 -5.35000000e-01 3.99999999e+00
9 262144 3.81469727e-06 1.95312500e-03 33 -5.00000000e-01 4.00000001e+00
10 1048576 9.53674316e-07 9.76562500e-04 2 -5.00000000e-01 4.00000001e+00
```



Beautiful tables: pgfplotstable (2/2)

level	DOF	e_1	e_2	info	∇e_2	$\frac{e_1^{(n)}}{e_1^{(n-1)}}$
1	4	2.5^{-1}	$7.58 \cdot 10^{-1}$	+48.0	–	–
2	16	6.3^{-2}	$5.00 \cdot 10^{-1}$	+25.0	–0.3	4
3	64	1.6^{-2}	$2.87 \cdot 10^{-1}$	+41.0	–0.4	4
4	256	3.9^{-3}	$1.44 \cdot 10^{-1}$	+8.0	–0.5	4
5	1 024	9.8^{-4}	$4.42 \cdot 10^{-2}$	+22.0	–0.85	4
6	4 096	2.4^{-4}	$1.70 \cdot 10^{-2}$	+46.0	–0.69	4
7	16 384	6.1^{-5}	$8.20 \cdot 10^{-3}$	+40.0	–0.52	4
8	65 536	1.5^{-5}	$3.91 \cdot 10^{-3}$	+48.0	–0.54	4
9	262 144	3.8^{-6}	$1.95 \cdot 10^{-3}$	+33.0	–0.5	4
10	1 048 576	9.5^{-7}	$9.77 \cdot 10^{-4}$	+2.0	–0.5	4



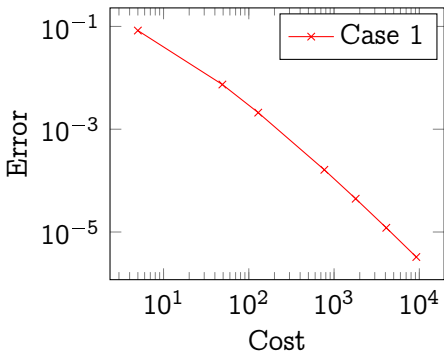
Choose the right chart for the job

- Take time to decide **which chart is appropriate**:
 - Which data points? What is the message?
 - Lines? Bars? Pie? Stacked bars? Surface? Or just a table?
 - Linear, categorical, or logarithmic axes?
 - Error bars?
- Present experimental results **honestly**, but in a way that highlight your contributions
- **Avoid** any form of **clutter** or visual gimmick, make it as readable as possible
- For articles, ensure readability when printed in black&white
- Consider using less common visualizations **when appropriate** (**radar chart**, **box plot**...). Be both **creative and readable**.



Beautiful plots: pgfplots

```
\begin{tikzpicture}
\begin{loglogaxis}[
xlabel=Cost,ylabel=Error]
\addplot[color=red,mark=x]
coordinates {
(5,8.31160034e-02)
(49,7.40715288e-03)
(129, 2.10192154e-03)
(769, 1.62269942e-04)
(1793, 4.44248889e-05)
(4097, 1.20714122e-05)
(9217, 3.26101452e-06)};
\legend{Case 1}
\end{loglogaxis}
\end{tikzpicture}
```





Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Online Presence

Do's

To Go Further

If it doesn't exist on the Internet, it doesn't exist.

Kenneth Goldsmith



Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Online Presence

Do's

To Go Further



Put preprints online

- As soon as a paper is done, put it online, in one of these forms:
 - On your Web page
 - As a technical report of your institution
 - As a preprint on arXiv/CoRR
- State very clearly in the paper and in references to it that it's an unpublished preprint
- Advantages:
 - People will know more easily about your work, will cite it, will reuse it
 - This can avoid getting scooped
 - Both science and your career will advance faster!
- Not in the following cases:
 - Double-blind review (though reviewers do not usually care)
 - Patent pending



Maintain published papers online

- Once your paper is accepted for publication, make sure it **remains online**
- Your Web site (if it is stable), arXiv, **publication repositories**, are suitable stores
- Same advantages as before!
- **Avoid** having **very different** versions online
- Sometimes publishers are not happy with this, but:
 - Some publishers directly allow you to do this (including ACM since recently); **prefer them!**
 - Many publishers allow an **almost-final** version to be put online
 - If you are a UK government employee or a US federal employee, you can always do this **by law**
 - **Open access** to science deserves a fight!



Put code and datasets online

- Computer science is special in that it should be easy to repeat experiments in exactly the same conditions; you just need to **share your code and datasets**
- Even if the code is not well-commented, better than nothing!
- You may want to keep it proprietary: but will it really be more used or more valuable than if you open-source it?
- Will ensure **reuse** of your research, further developments
- Clear **advantage** during the reviewing process

Example

RoadRunner (Creszenzi et al., 2001): early work on Web data extraction. Not the first, not the best, many refinements since. Remains the most cited reference in the area (1024 citations!) because **the system (and datasets) is freely available online.**



Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Online Presence

To Go Further



References

The Chicago Manual of Style. The University of Chicago Press, 16th edition, 2010.

William S. Cleveland. *The Elements of Graphing Data*. Hobart Press, revised edition, 1994.

Stephen Few. *Show me the Numbers*. Analytics Pres, 2004.

Edward Tufte. *The Visual Display of Quantitative Information*. Graphics Press LLC, 2001.

Edward Tufte. *Beautiful Evidence*. Graphics Press LLC, 2006.

Mary-Claire van Leunen. *A Handbook for Scholars*. Oxford University Press, revised edition, 1992.

Merci.