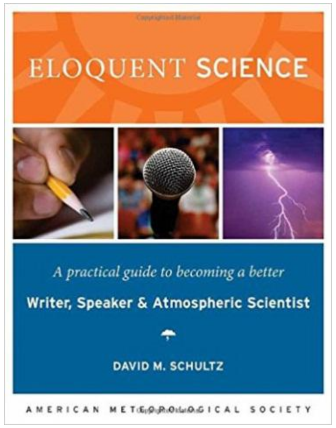
Good (and bad) practices for effective scientific illustration

Tzung-May Fu (tmfu@pku.edu.cn)
Dept. of Atmospheric & Oceanic Sciences, Peking University

Atmospheric Chemistry & Climate @ Peking University
https://www.phy.pku.edu.cn/atmoschem

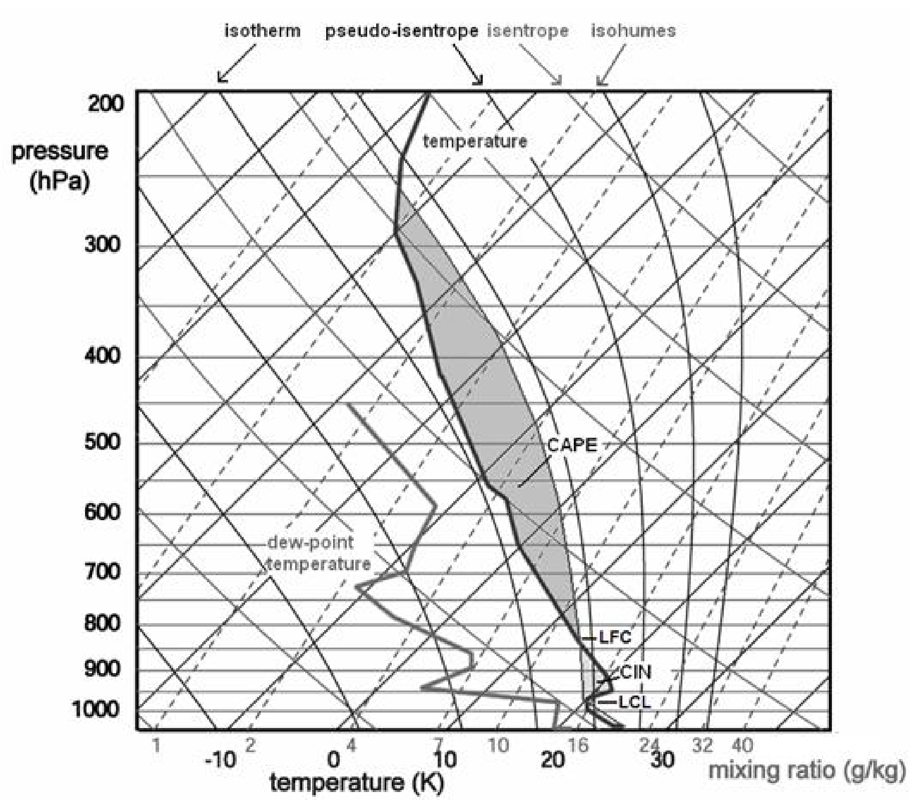
Resources

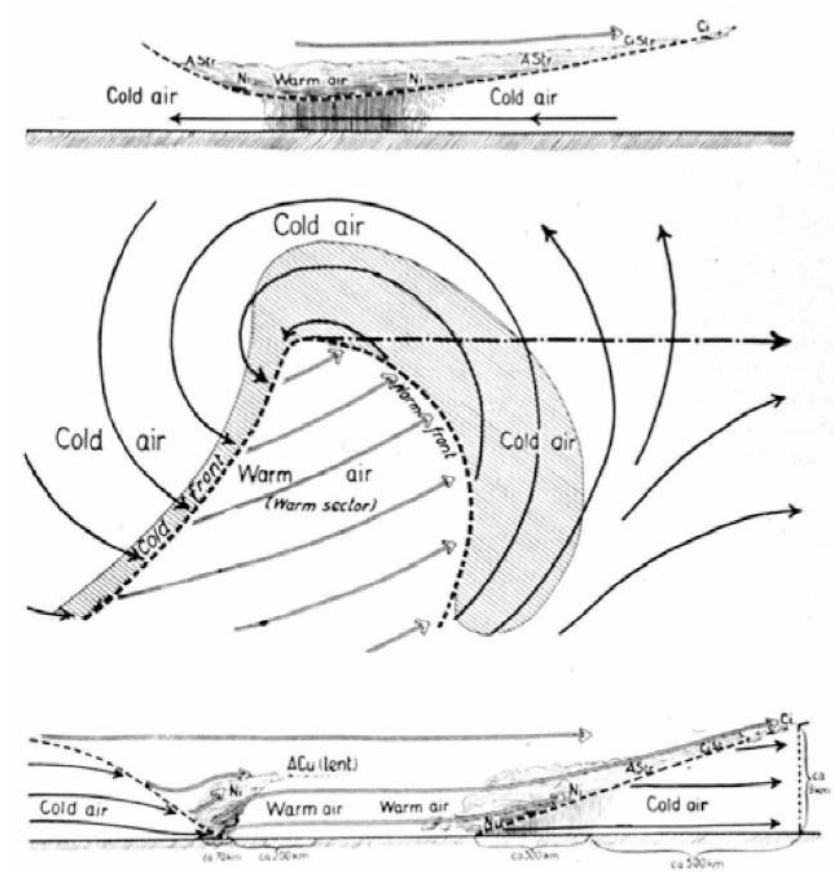


- **David Schultz's book *Eloquent science* and his blog eloquentscience.com/**
- **Buckingham (2008), Effective scientific illustrations, *Lab Times***
- **Rolandi et al. (2011), A Brief Guide to Designing Effective Figures for the Scientific Paper, *Advanced Materials***
- **Rougier et al. (2014), Ten simple rules for better figures, *PLOS Computational Biology***
- **PSU College of Earth & Mineral Sciences, Style for students**

“Data figures” and “concept figures”

We will mainly focus on “data figures” prepared for scientific publications and presentations





[Bjerknes, 1919]

Why do we plot?

"Scientific visualization is a graphical interface between people and data."

Rougier et al. (2014)

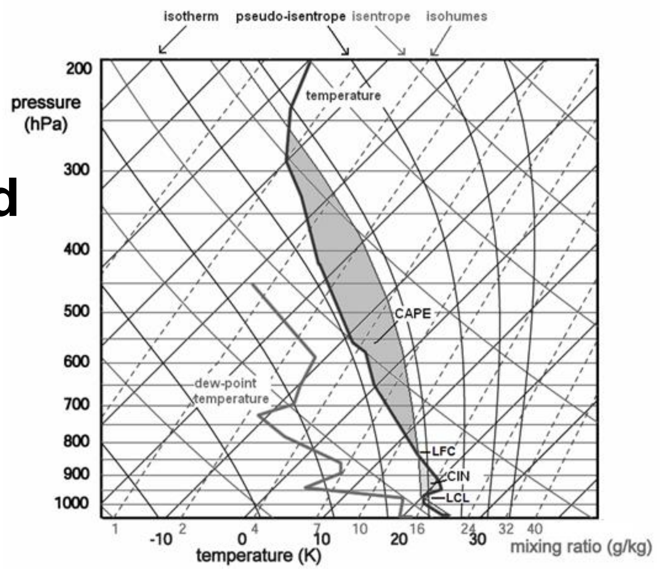
- **To provide more information:** to visualize ideas that are hard to show otherwise
- **To focus attention:** to direct focus to ideas that otherwise may be easily buried in texts and tables
- **To tell a story:** to build connections between plot elements or between a series of plots
- **To broadcast our work:** figures travel faster and further than texts

"You should always have one memorable figure in your paper."

~ Daniel Jacob

Key points to consider and revisit

1. **Know your audience.**
2. **Know your key message.** Readers should be able to get the main message without reading the text.
3. **Be honest.** Use real data and show uncertainty. Do not mislead readers.
4. Show the data, not the design.
5. Assess necessity.
6. Throw away plots that don’t work.



Why not Excel?

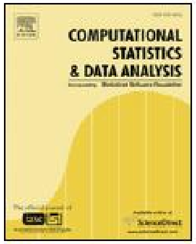
- **Excel is ok for cursory data exploration to produce "working figures", but it does not produce publication-quality figures**
- **Matlab, IDL, NCL take effort but are worthwhile in the long run**

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

ELSEVIER



It's easy to produce chartjunk using Microsoft®Excel 2007 but hard to make good graphs

Yu-Sung Su

Department of Political Science, The Graduate Center, The City University of New York, 365 Fifth Avenue, New York, NY 10016, USA

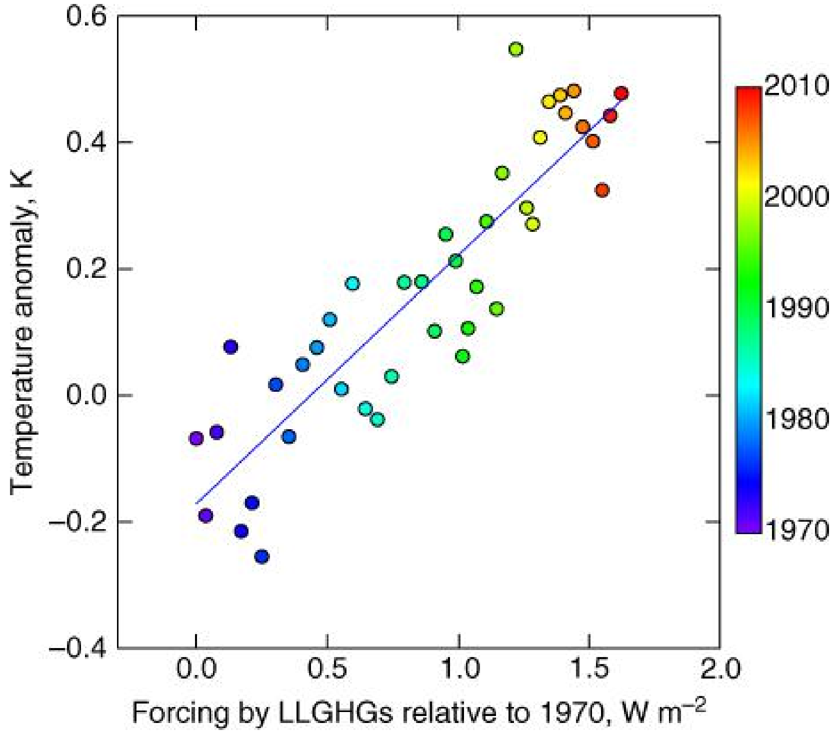
10 basics for making effective plots

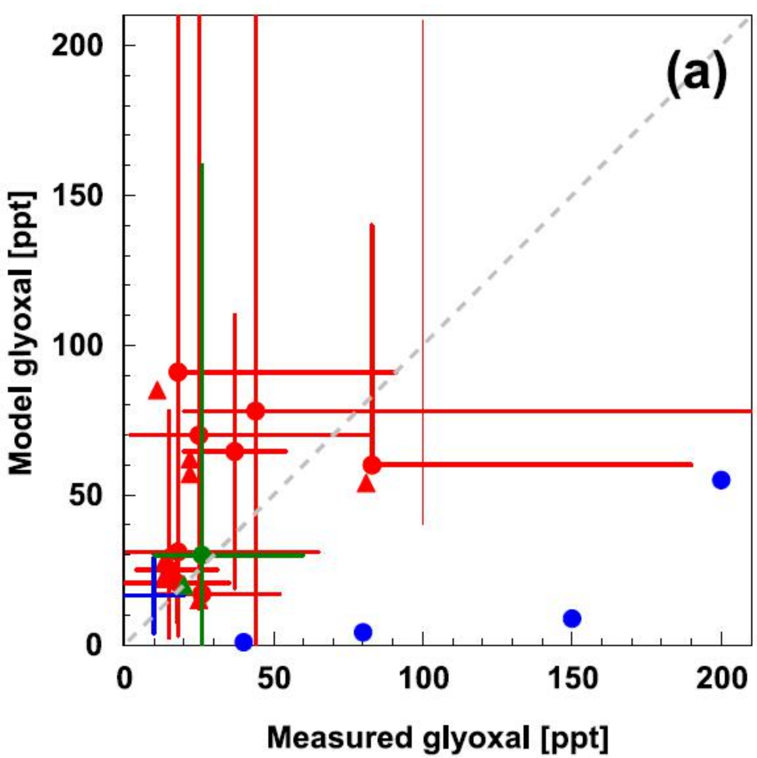
1. Choose the right type of plot. No pies!
2. Consider size, ratios, axes, directions
3. Optimize lines & symbols.
4. Optimize colors & color scales.
5. Optimize labels & legends.
6. Use consistent fonts and font sizes.
7. Use 3-D effects and animations with caution.
8. Focus on key message and simplify. Maximize data-to-ink ratio. Adapt to the media.
9. Write full captions from the very beginning.
10. Choose the right format and resolution.

1. Choose the right type of plot (No pies!)

Choose the simplest type of plot that does the job!

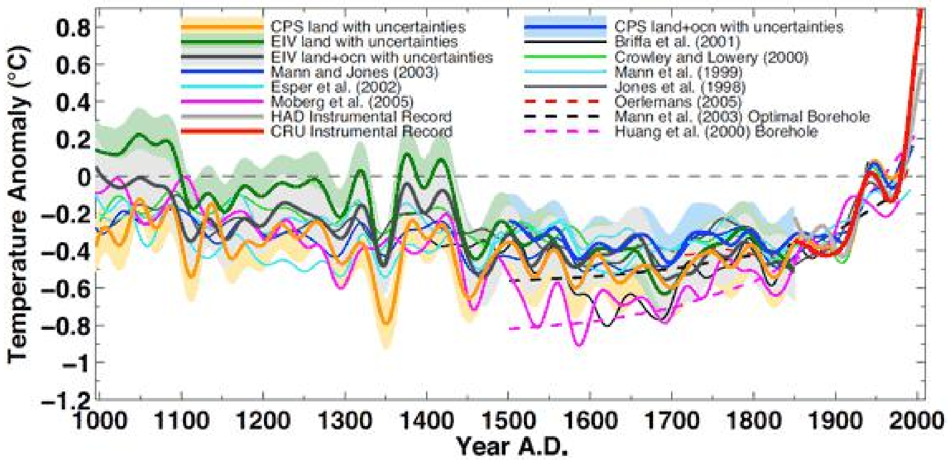
- **Scatter plot: show correlation between 2 numerical variables (x: predictor variable, y: response variable) or (x: observations, y: model)**

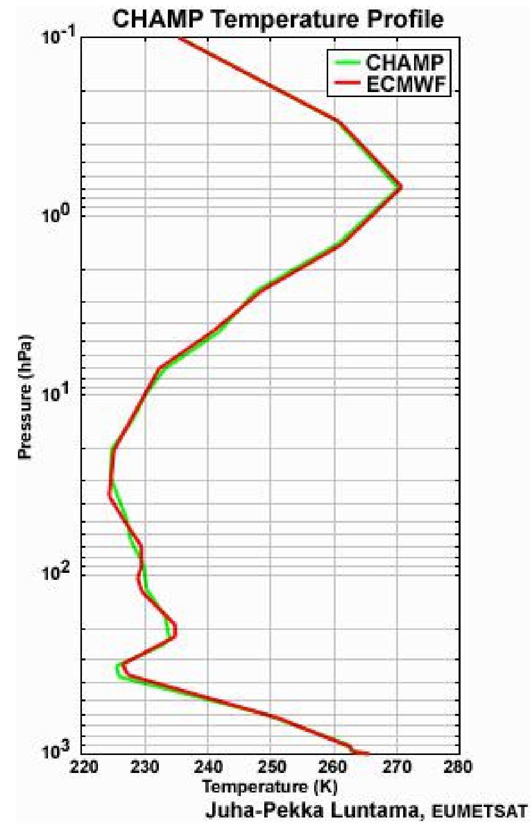


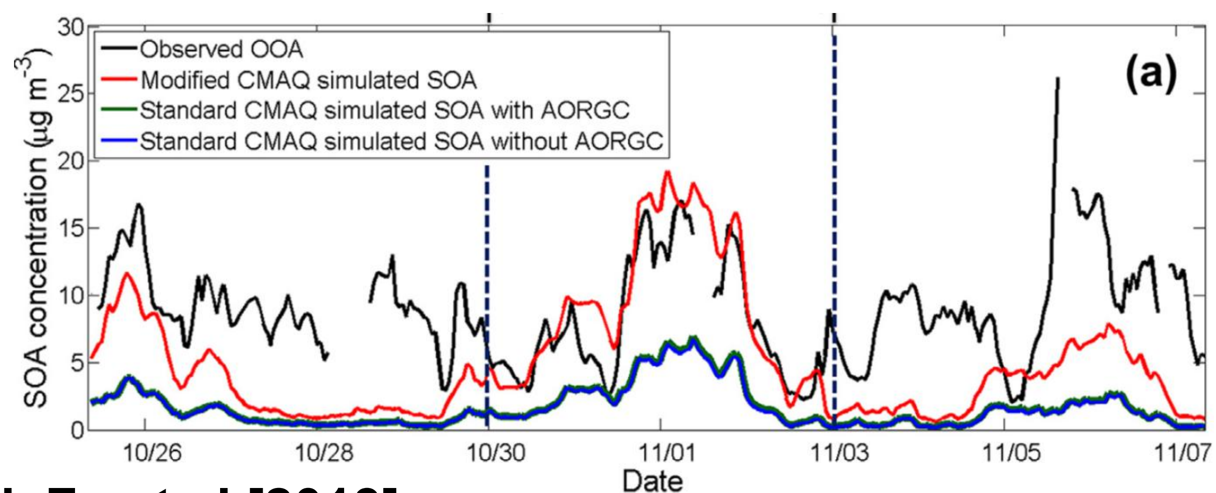


1. Choose the right type of plot (No pies!)

➢ **Line plot: show continuous changes in a numerical variable (timeseries, profiles)**



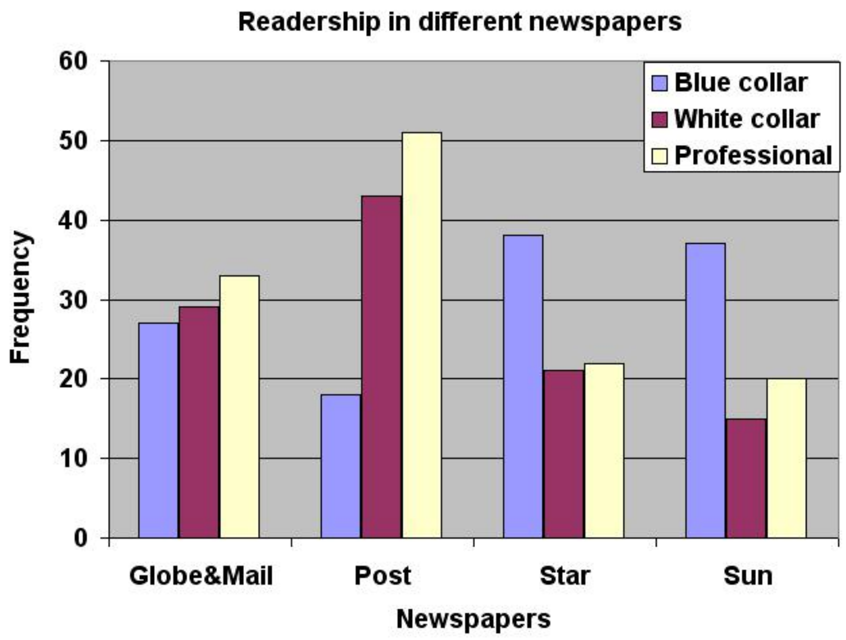


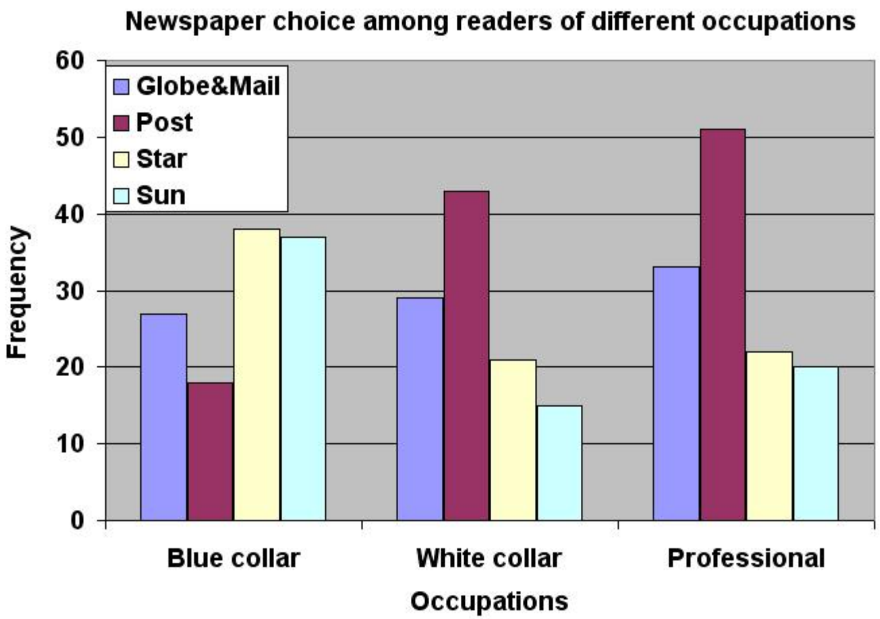


Li, Fu et al [2013]

1. Choose the right type of plot (No pies!)

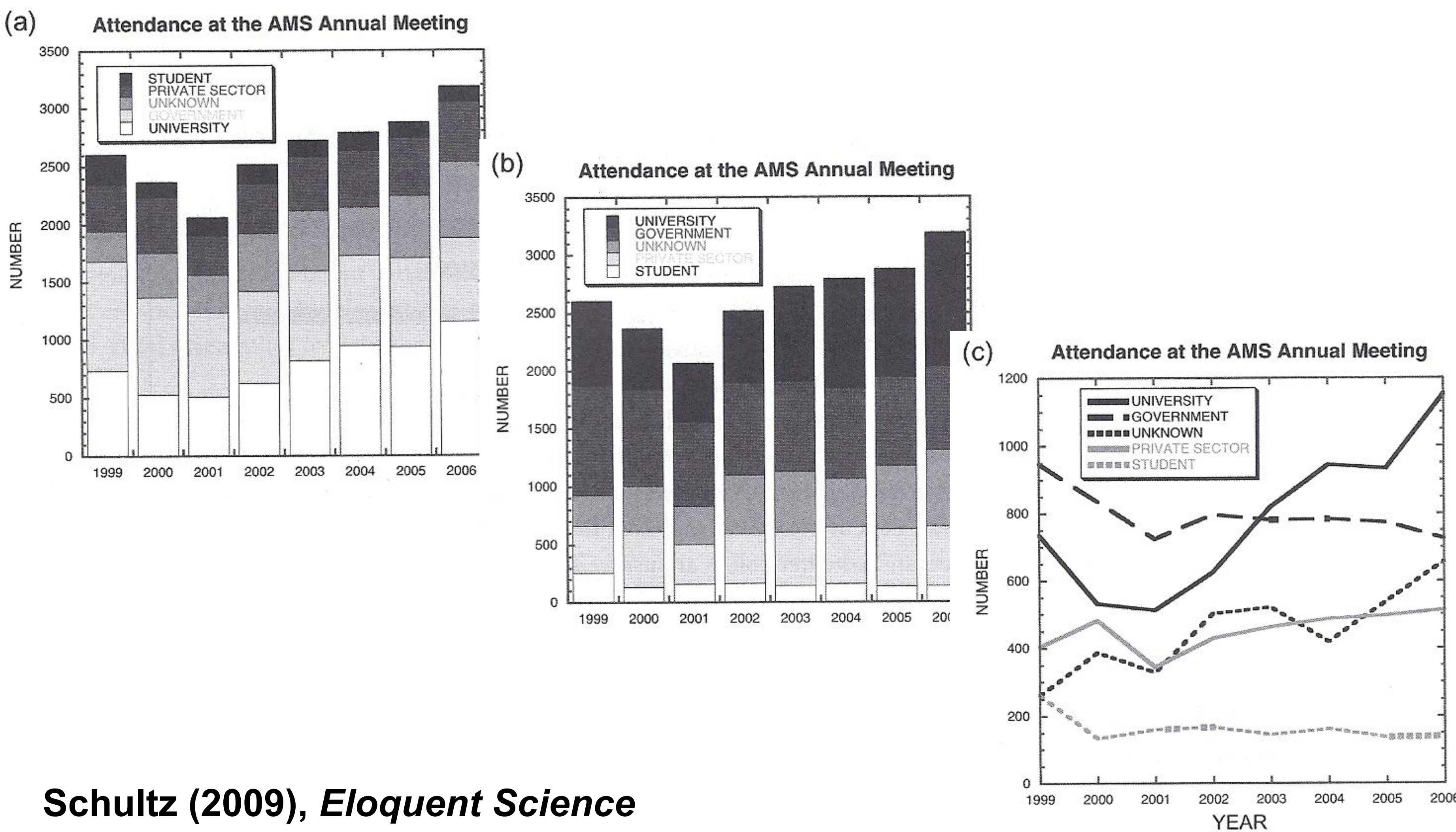
- **Bar chart: good for comparing differences between categorical variables**





1. Choose the right type of plot (No pies!)

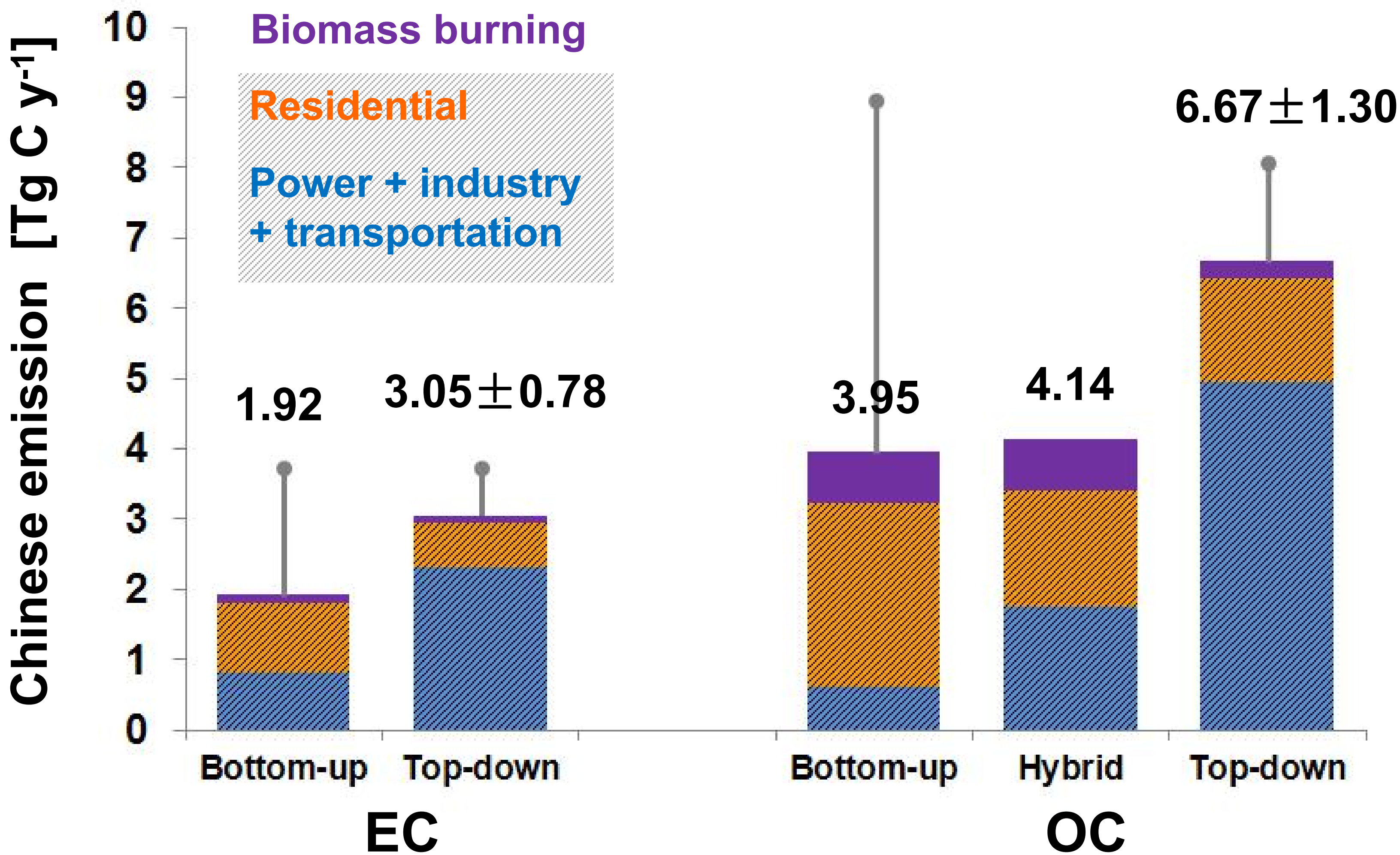
Line charts and bar charts can sometimes better alternatives for each other



Schultz (2009), *Eloquent Science*

1. Choose the right type of plot (No pies!)

- Pie chart: can usually be replaced by bar charts

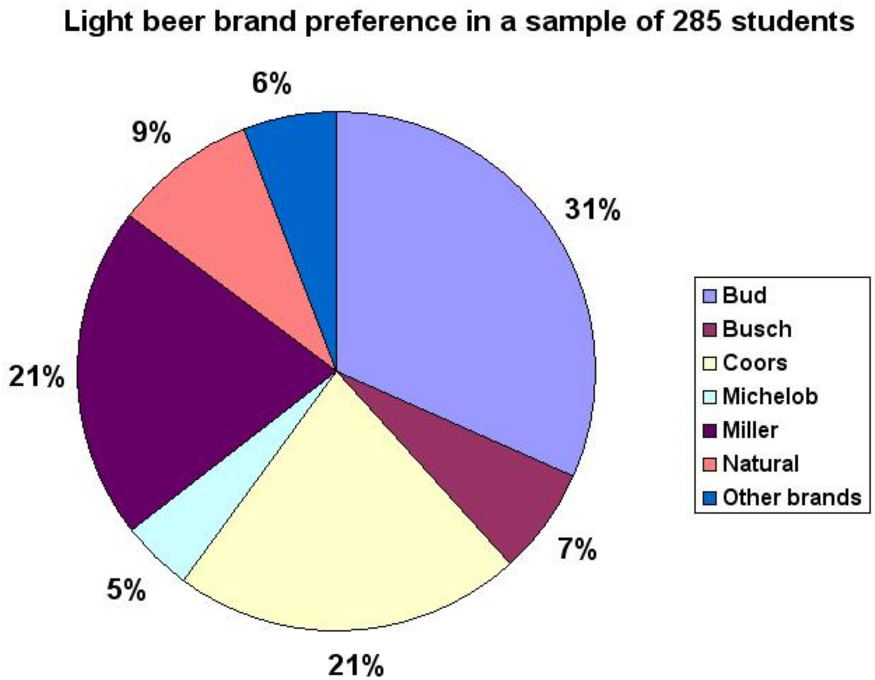


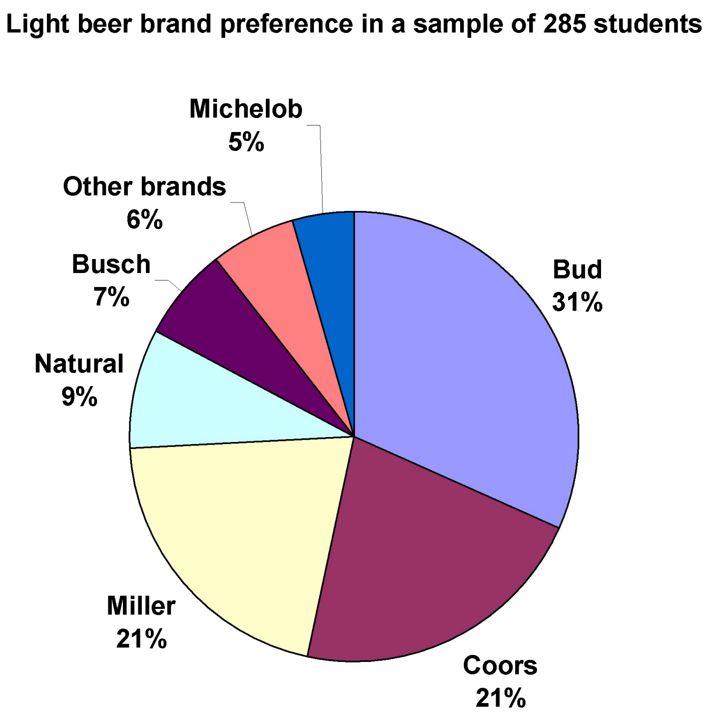
Fu et al. [2012]

1. Choose the right type of plot (No pies!)

- Pie chart: can usually be replaced by bar charts

If you absolutely must use pies …

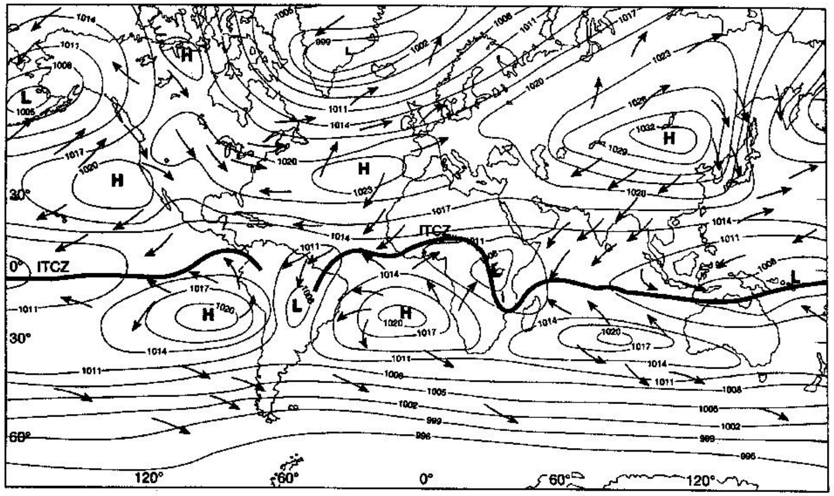




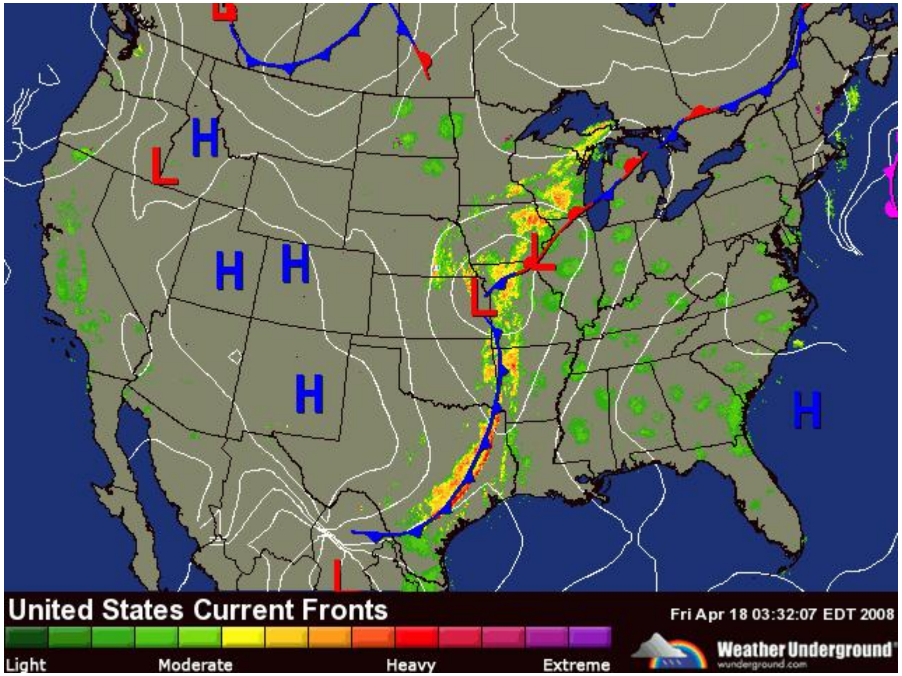
Which one is better? Why?

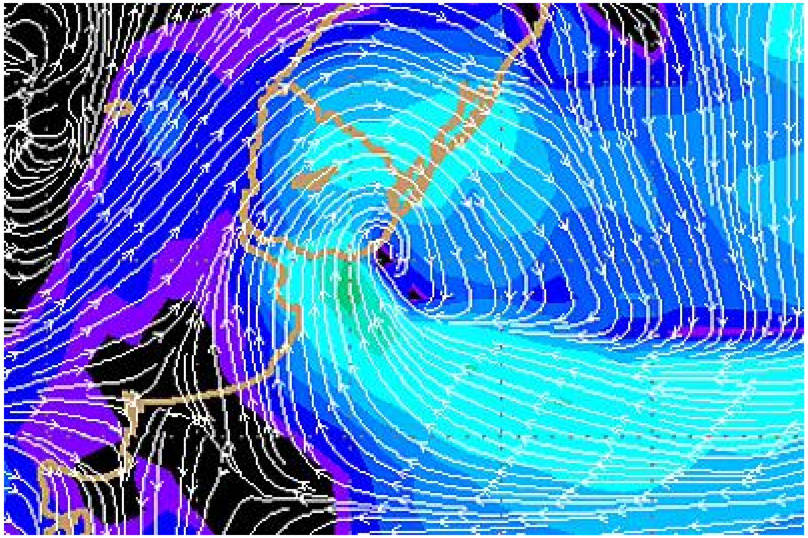
1. Choose the right type of plot (No pies!)

- Maps: geographical locations
- Vector maps
- Streamline maps
- Weather maps



(a) January

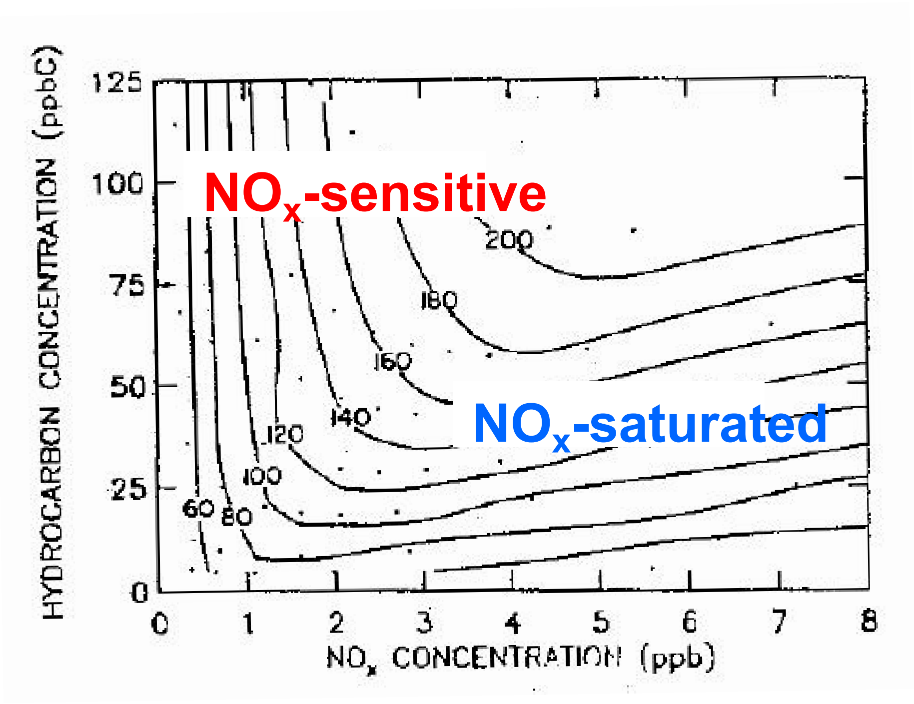




1. Choose the right type of plot (No pies!)

- Contour plots: show continuous changes in 2-D space

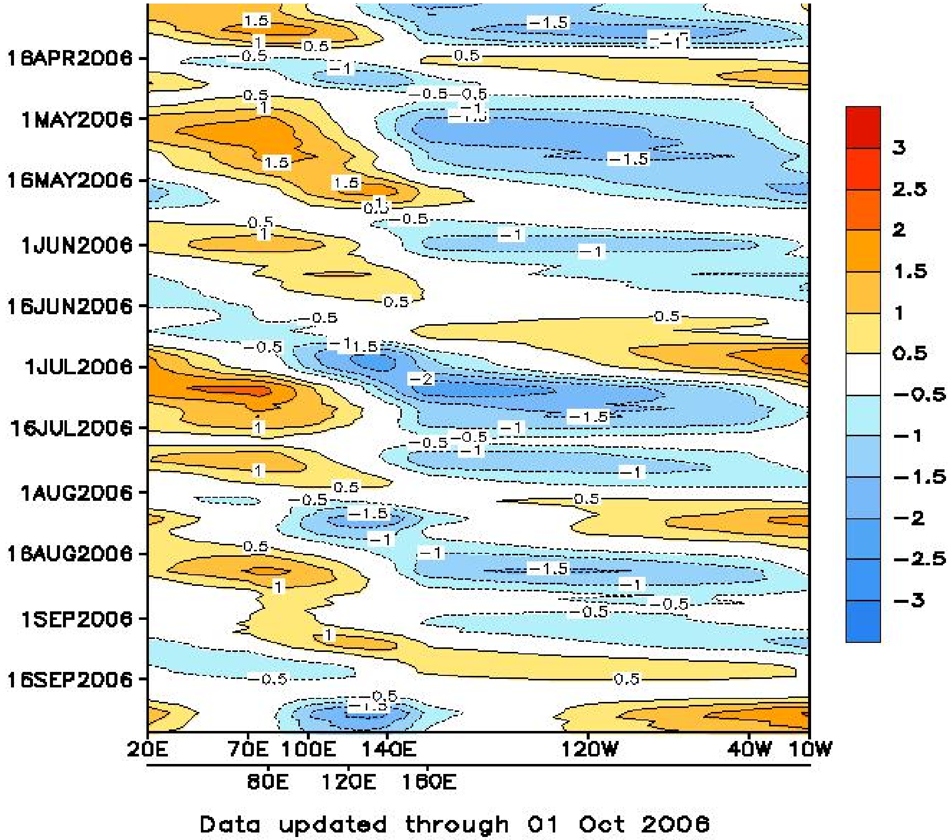
Surface O_3 concentration



Sillman et al. [1990]

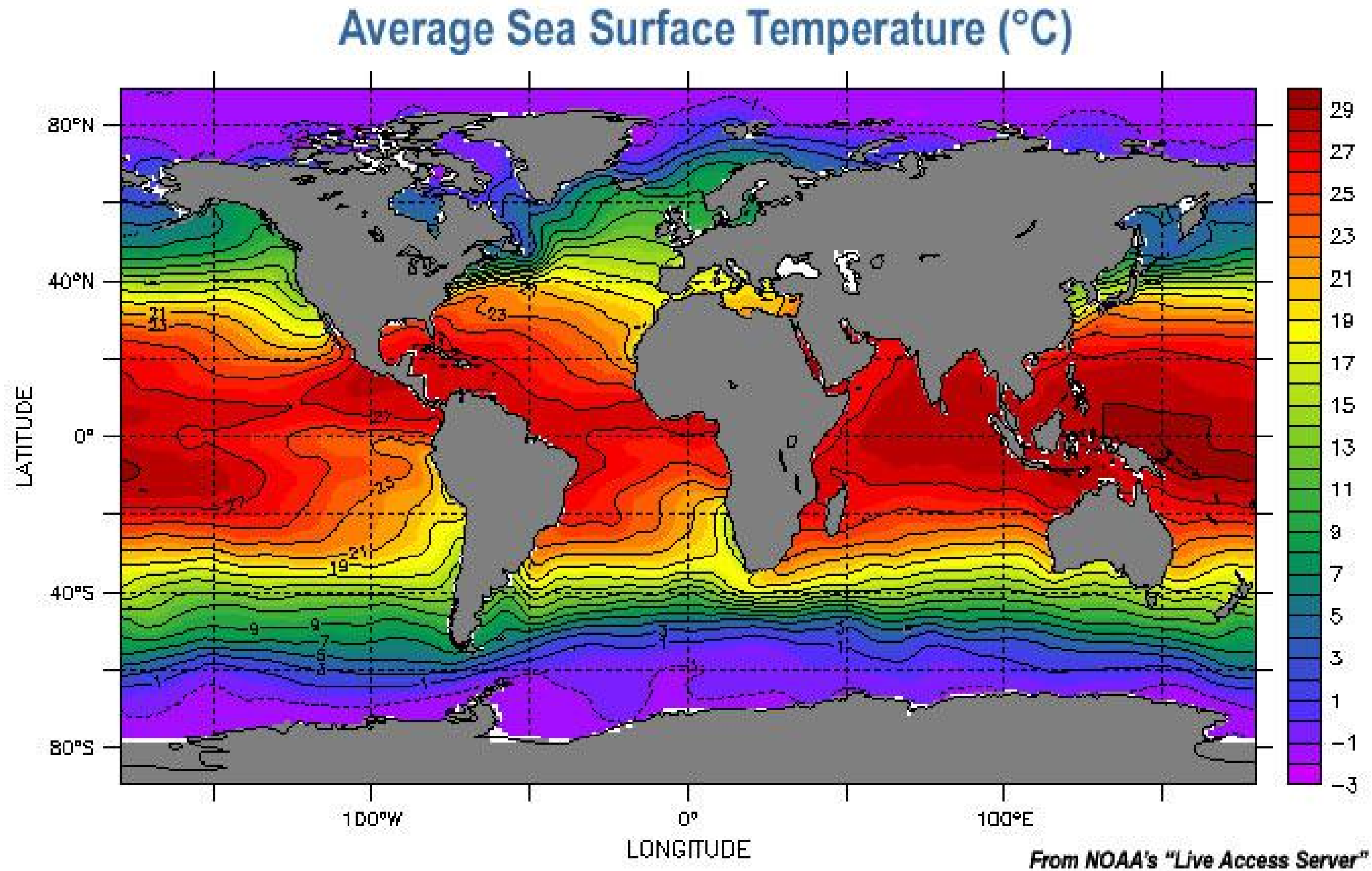
Hovmöller plot

5-day running mean OLR anomaly



1. Choose the right type of plot (No pies!)

- **Contour maps: show continuous change on a map**

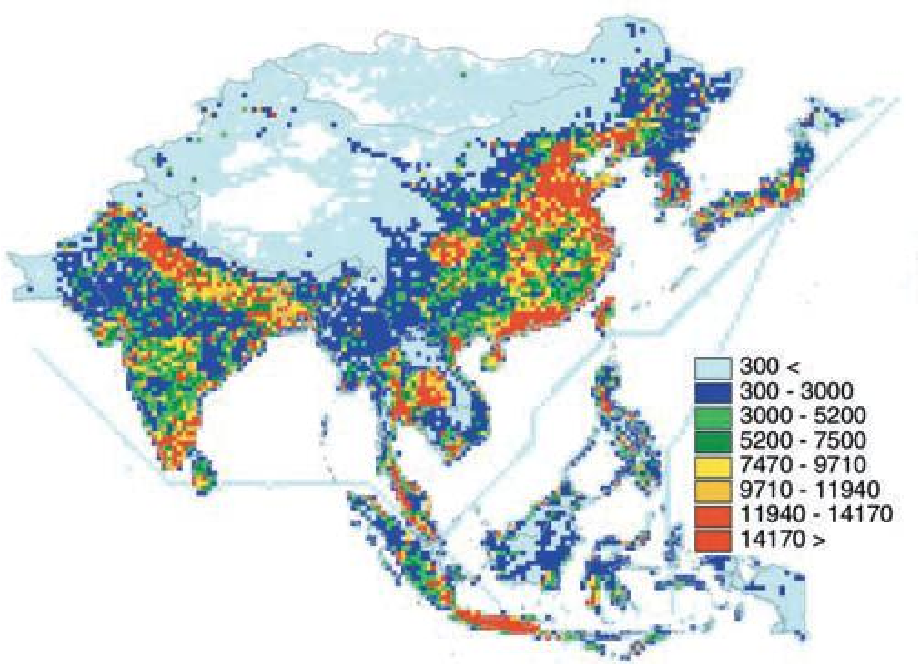


From NOAA's "Live Access Server"
Levitus 1982 Annual Climatology

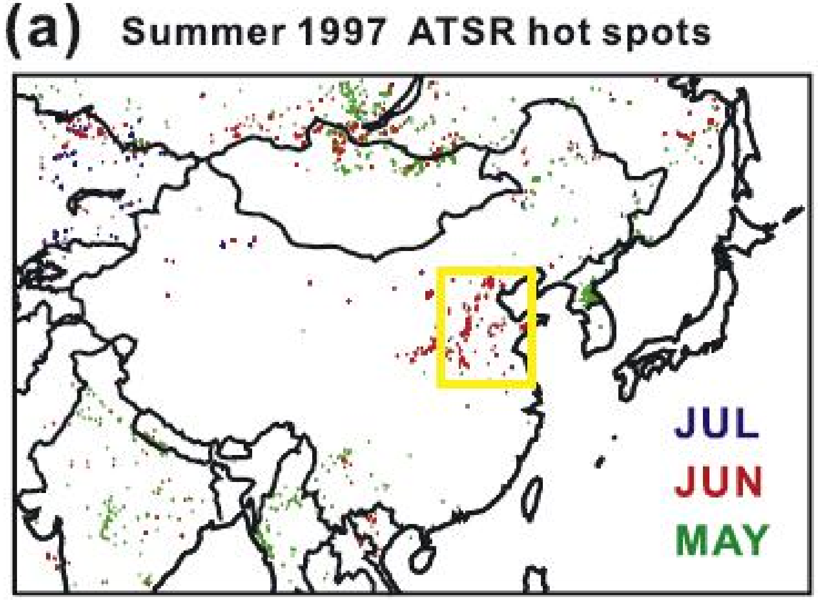
1. Choose the right type of plot (No pies!)

- Pixelated maps: show discrete changes over a map and emphasize hotspots

Asian NMVOC emissions for 2001
[Mg y^{-1} box^{-1}]



Streets et al. [2003]

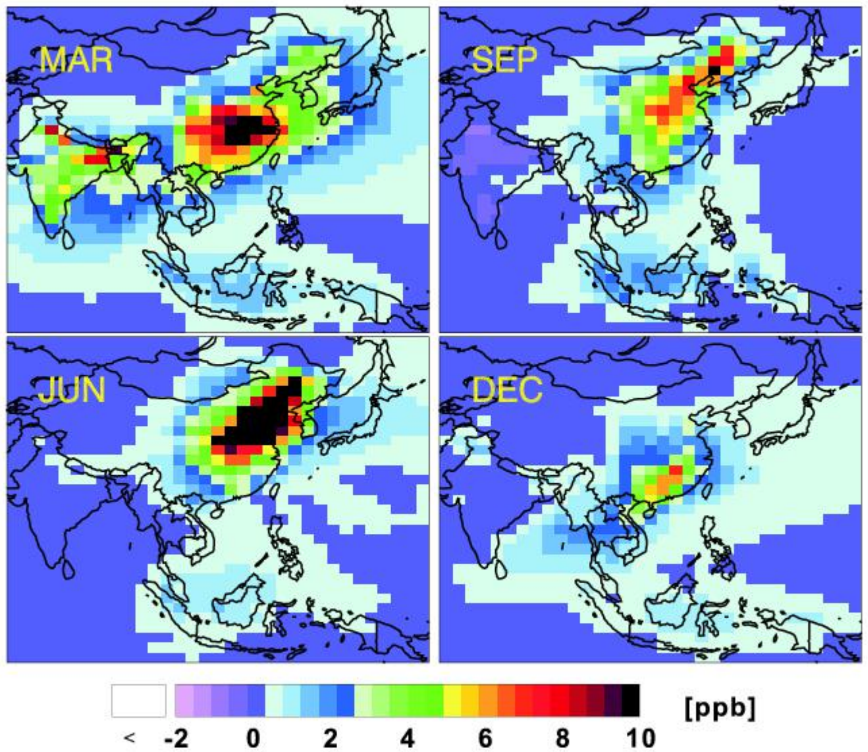


Fu et al. [2007]

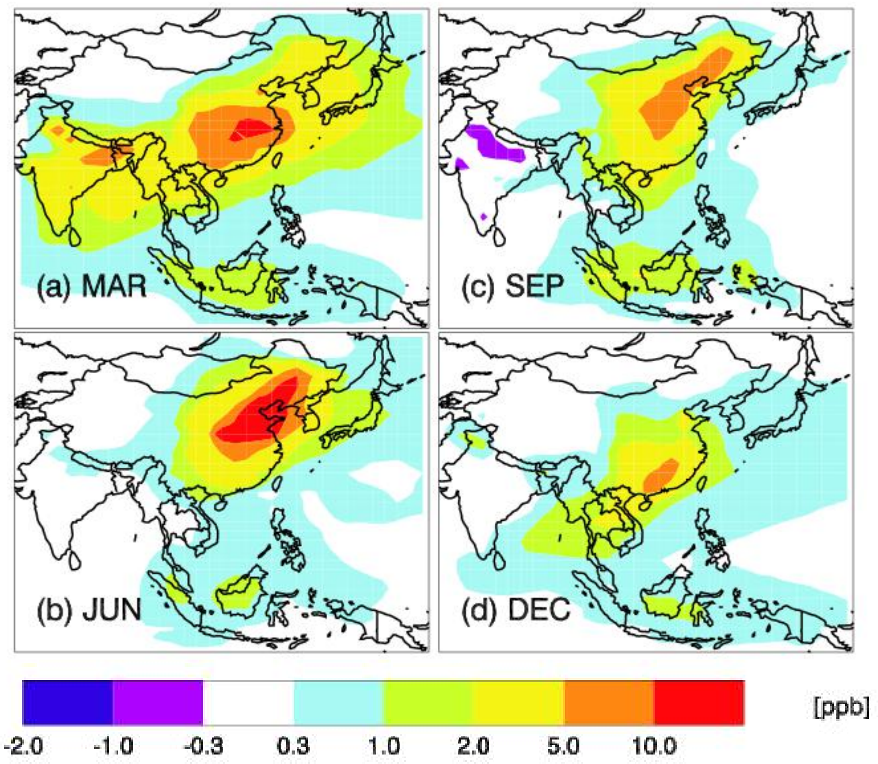
1. Choose the right type of plot (No pies!)

Simulated ΔO_3 using two different emission inventories

Pixelated map



Filled contour map

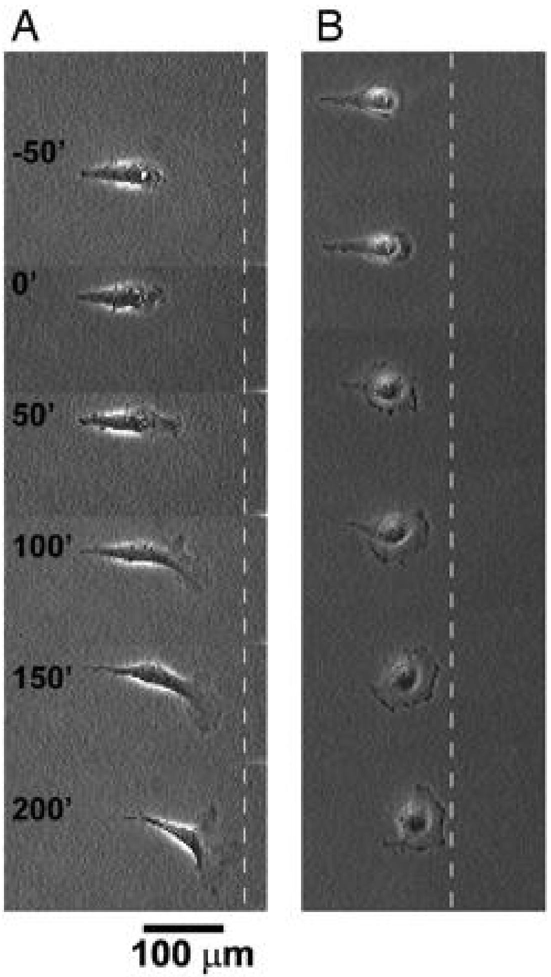


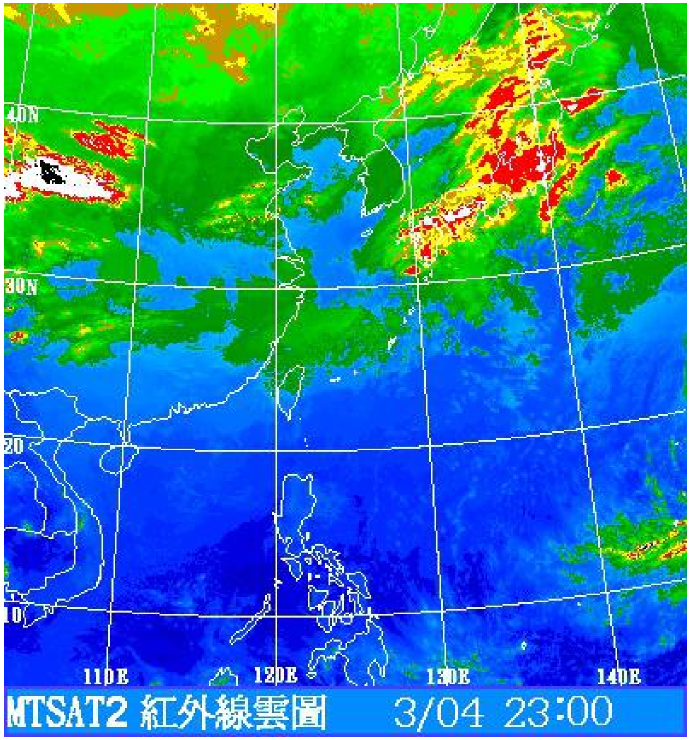
Which one is more appropriate?

Fu et al. [2007]

1. Choose the right type of plot (No pies!)

- **Photos: visual results from experiments and observations**



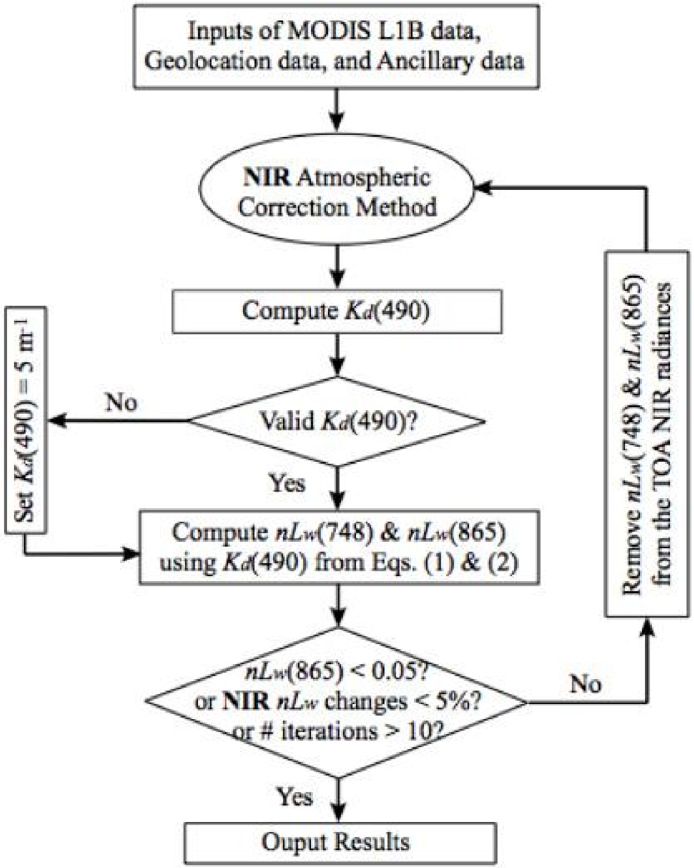


Jiang et al. [2004]

1. Choose the right type of plot (No pies!)

Other types of plots:

- **Flow charts: protocols, algorithms**



- **Conceptual plots: processes**

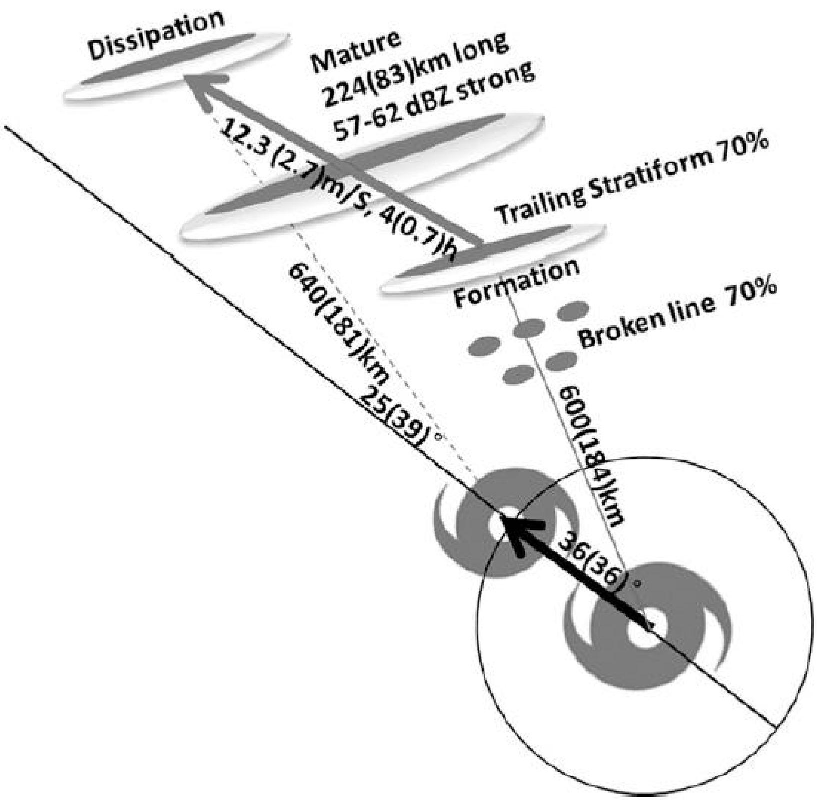


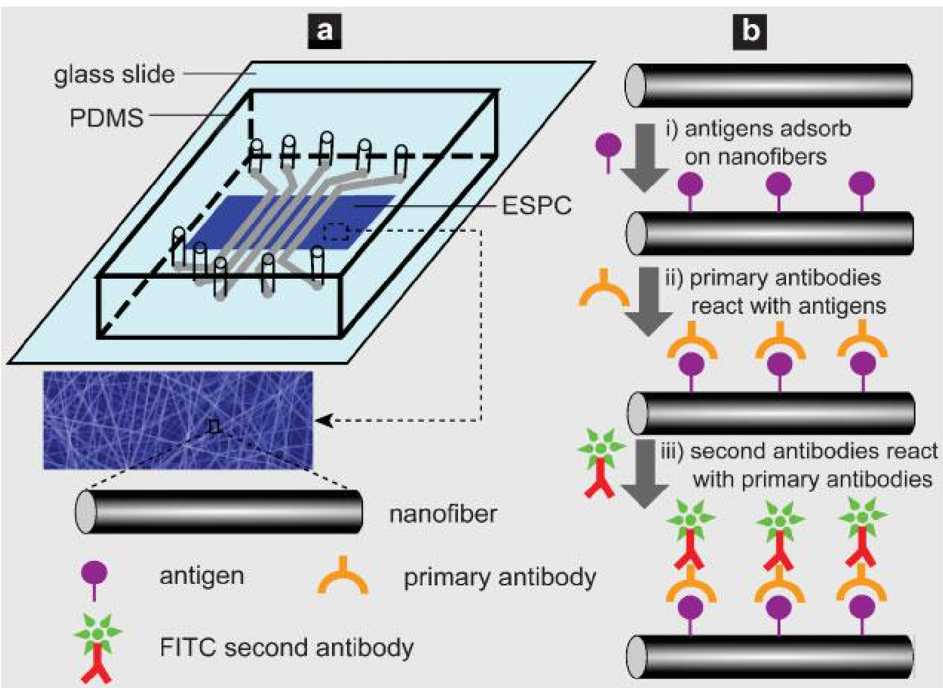
FIG. 5. Schematic diagram of the averaged features of pre-TC squall lines. The number in parentheses is the corresponding standard deviation.

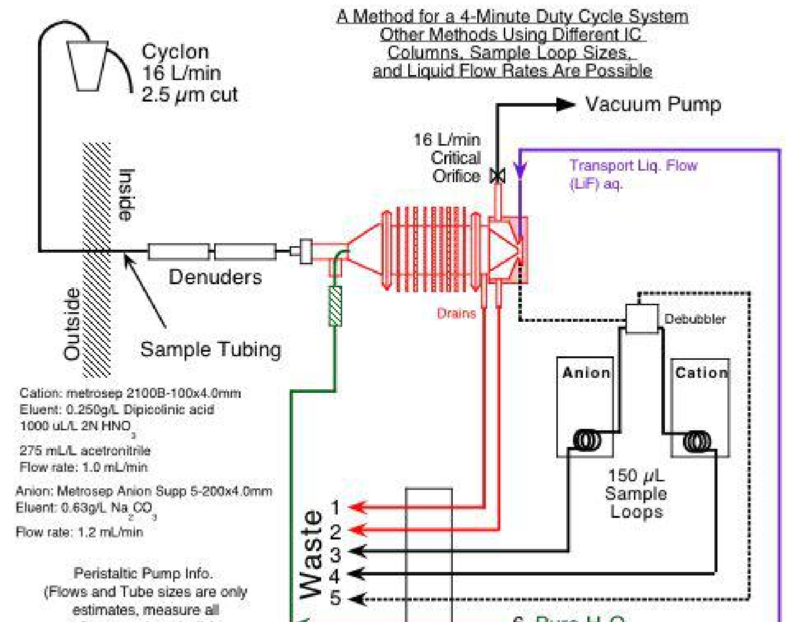
[Meng et al., 2013]

1. Choose the right type of plot (No pies!)

Other types of plots:

- Apparatus / methodology





Tubing	Flow, mL/min	Colour
1	-	green/green
2	-	green/green
3	0.06	orange/blue
4	0.06	orange/blue
5	adjust	orange/white
6	1.4 to 1.6	yellow/yellow
7	0.17 to 0.2	orange/green

Not everything needs to be a plot!

Table 4. Global Sources of Secondary Organic Aerosol in GEOS-Chem[a]

Mechanism	Reversible	Reversible	Reversible	Irreversible	
Precursors	Terpenes, Terpenoid Alcohols, and Sesquiterpenes	Isoprene	Aromatics[b]	Glyoxal and Methylglyoxal	Total
Source, Tg C a^{-1}	9.0	6.7	1.7	11[c]	29
Above 1.5 km	3.1	2.8	0.95	4.4	
Below 1.5 km	5.9	3.9	0.72	6.2	
Annual mean burden, Tg C	0.19	0.21	0.04	0.17[d]	0.62
Above 1.5 km	0.14	0.18	0.03	0.12	
Below 1.5 km	0.05	0.03	0.01	0.05	

[a]Including reversible formation from condensation of semivolatile products of VOC oxidation, and irreversible formation from oxidation and/or oligomerization of glyoxal and methylglyoxal in aqueous aerosols and clouds. Global production rates of each aerosol type formed by the reversible pathway are derived from balance with deposition. The organic to carbon mass ratio for all reversible-partitioning SOA is assumed to be 2.1, following *Henze et al.* [2008].

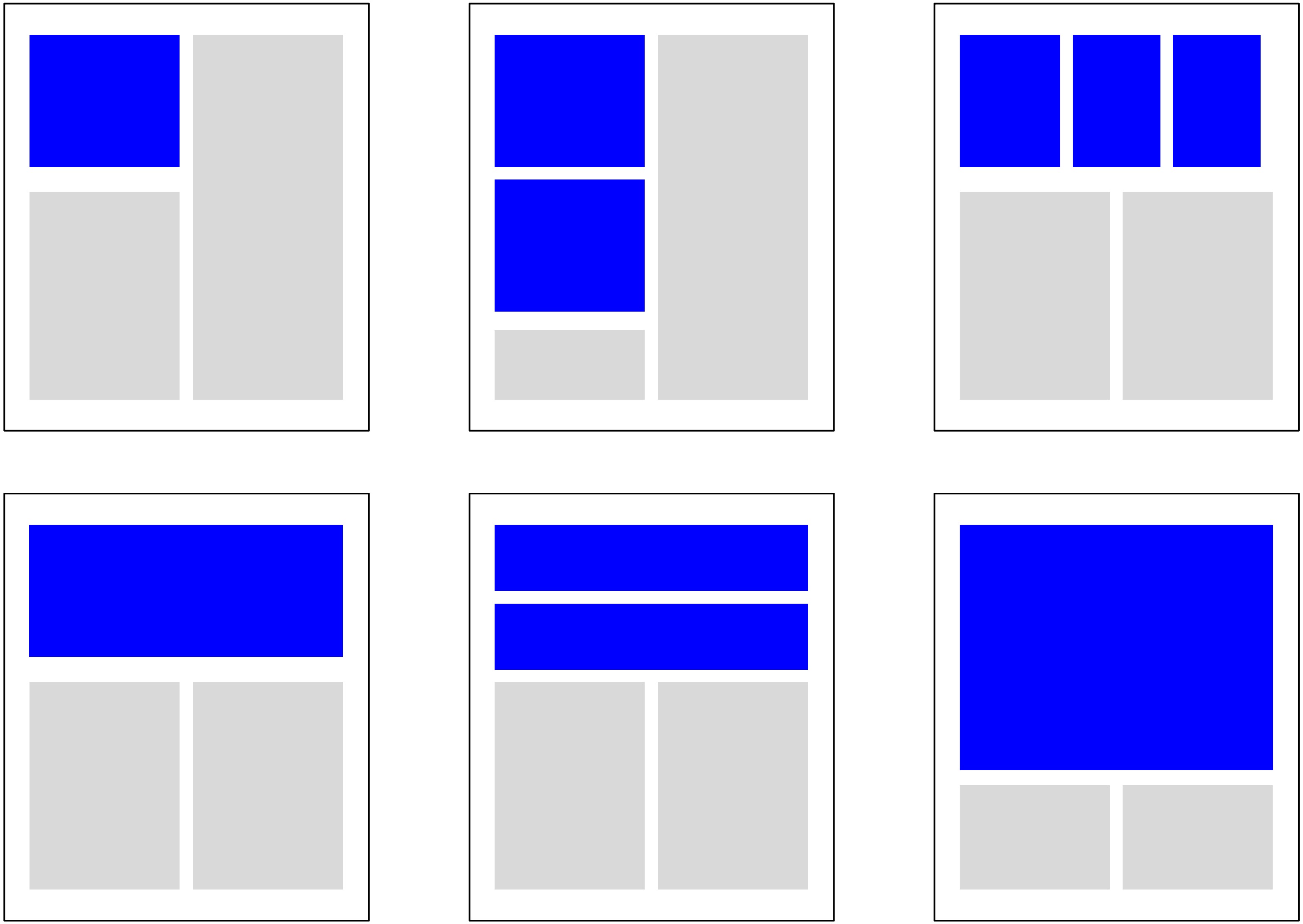
[b]From *Henze et al.* [2008].

[c]Includes 2.6 Tg C a^{-1} from glyoxal and 8 Tg C a^{-1} from methylglyoxal (Table 2).

[d]Includes 0.04 Tg C from glyoxal and 0.13 Tg C from methylglyoxal (Table 2).

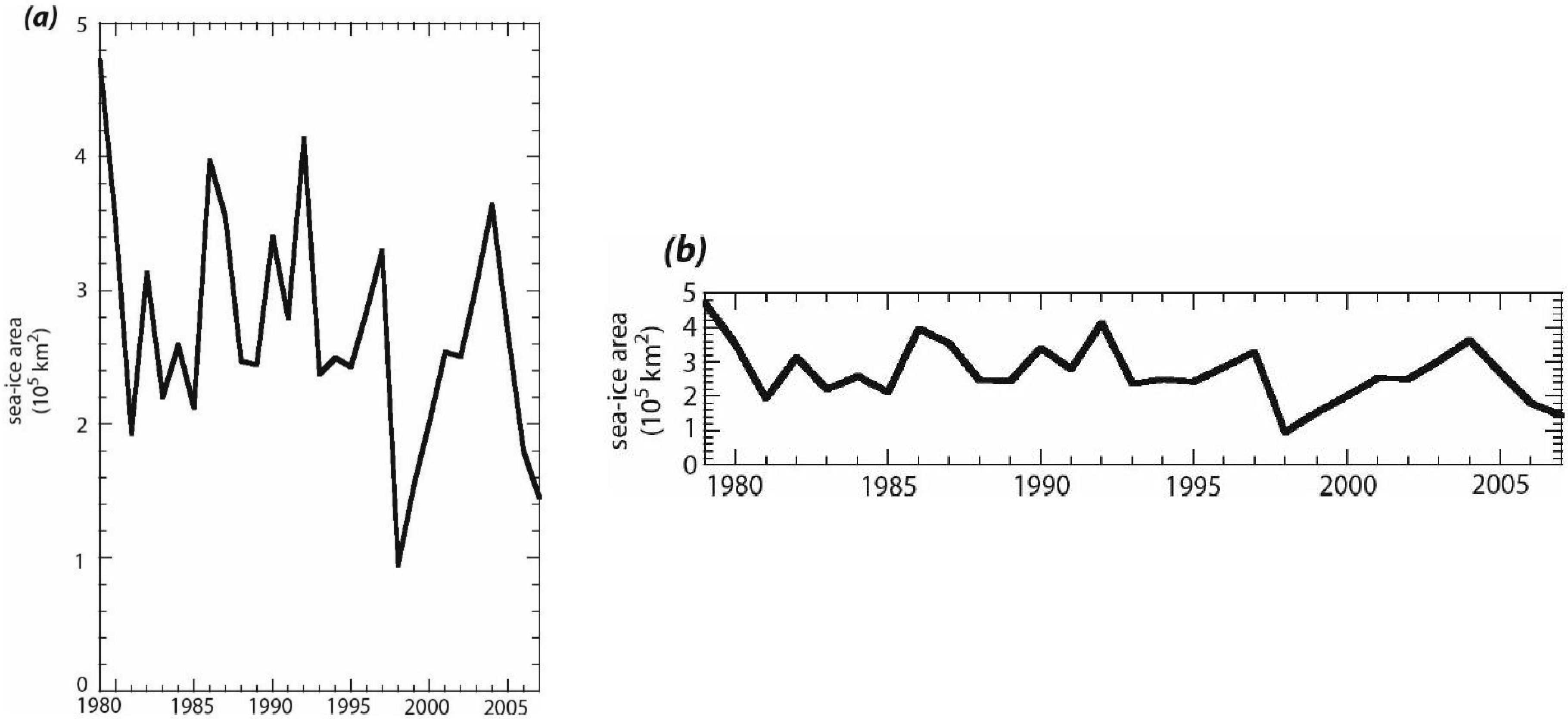
Fu et al. [2008]

2. Consider size, ratios, & axes



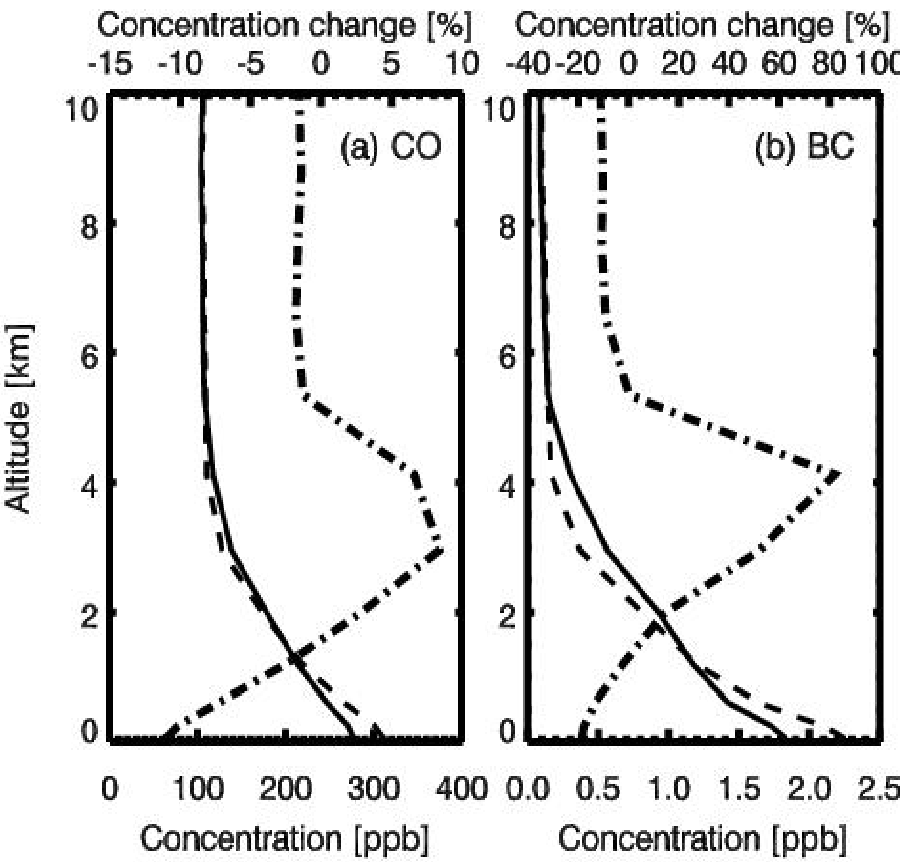
2. Consider size, ratios, & axes

During which period did the sea ice area experience the sharpest decline?

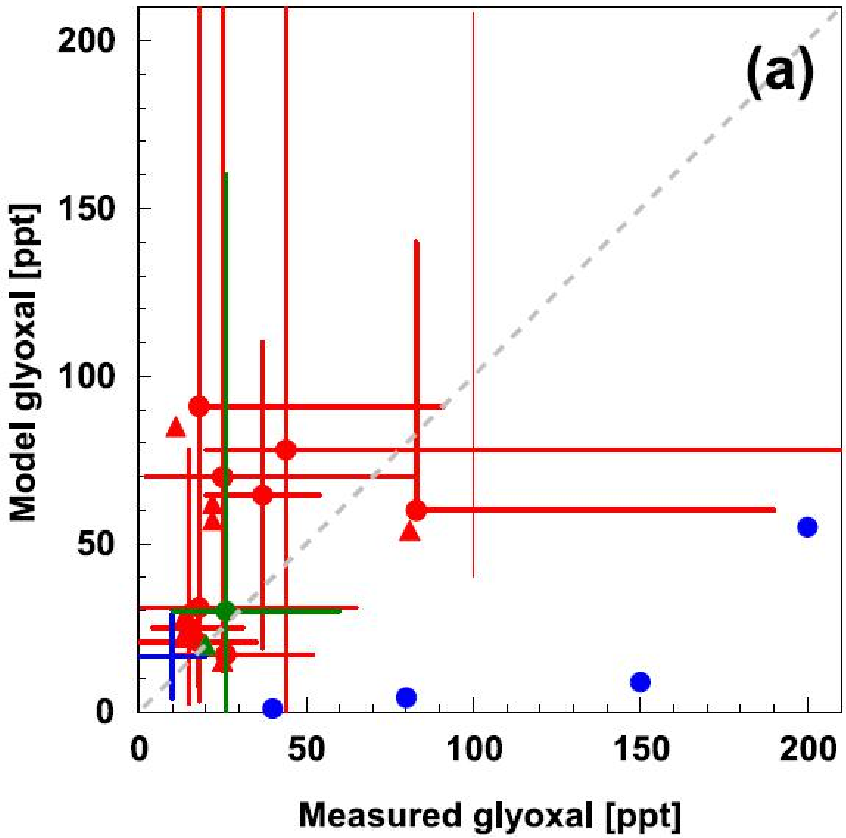


Schultz (2009), *Eloquent Science*

2. Consider size, ratios, & axes



Jian and Fu [2014]

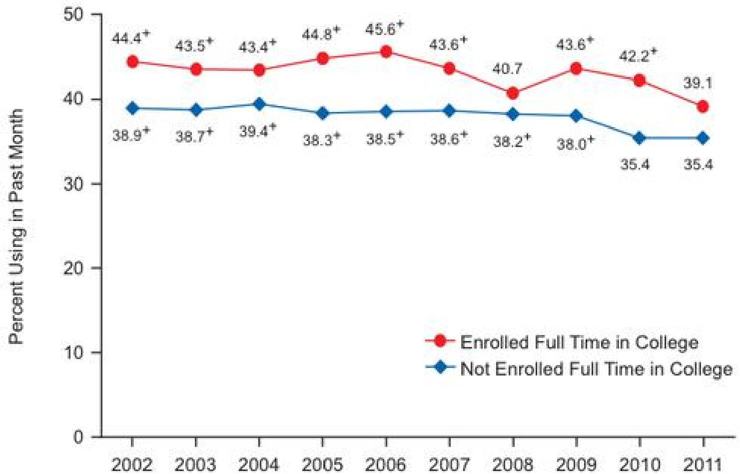


Fu et al. [2008]

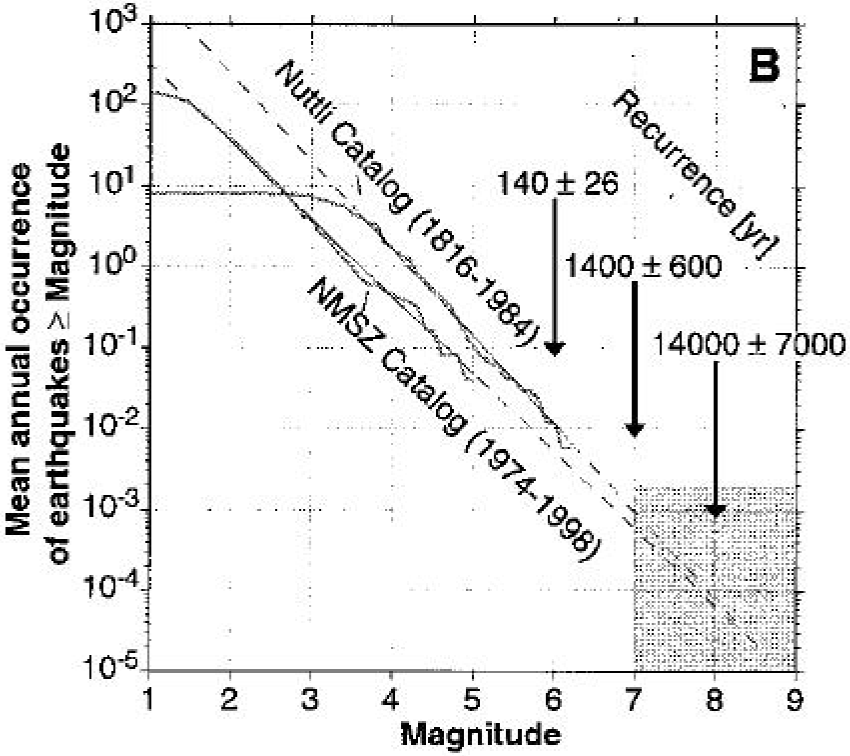
2. Consider size, ratios, & axes

Choose appropriate axis style and range

Figure 3.3 Binge Alcohol Use among Adults Aged 18 to 22, by College Enrollment: 2002-2011

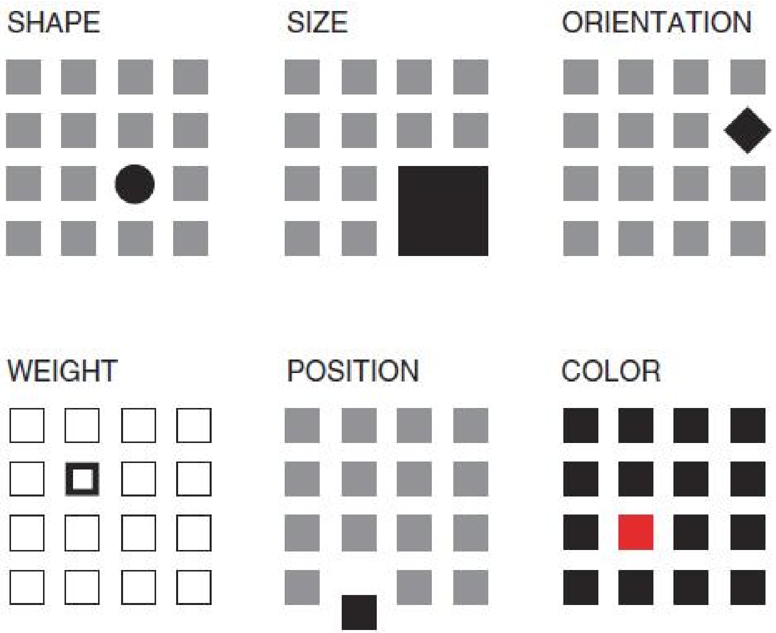


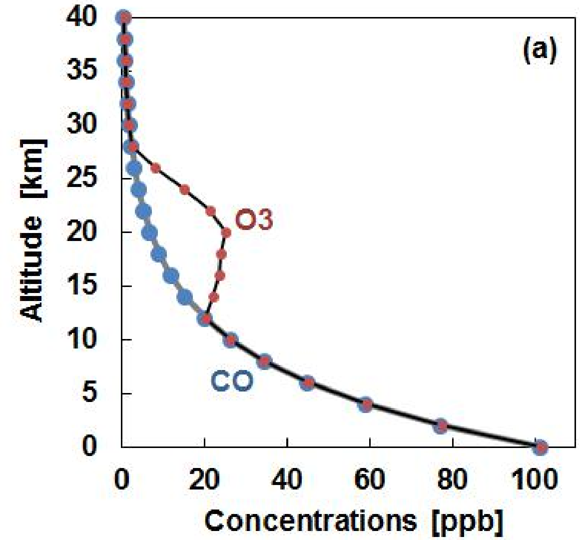
+ Difference between this estimate and the 2011 estimate is statistically significant at the .05 level.

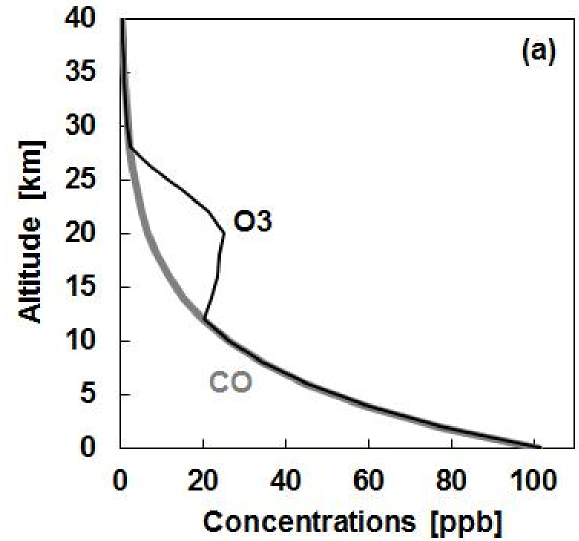


3. Optimize lines and symbols

There are many types of contrast, but one (at most two) is usually enough



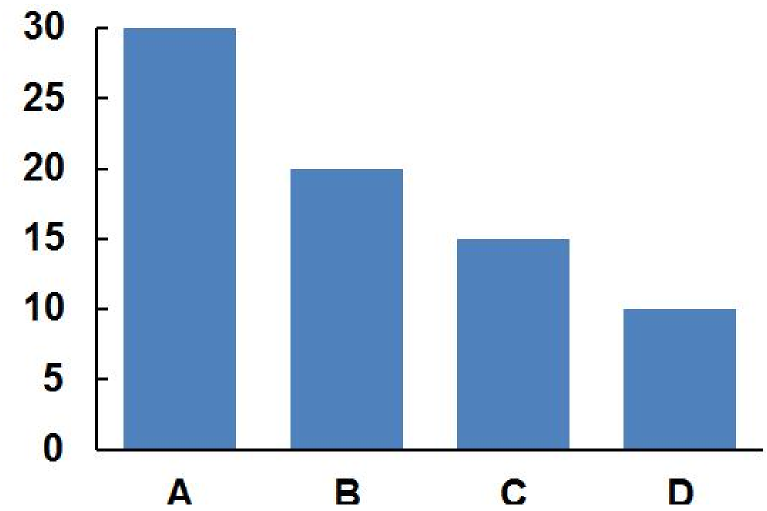


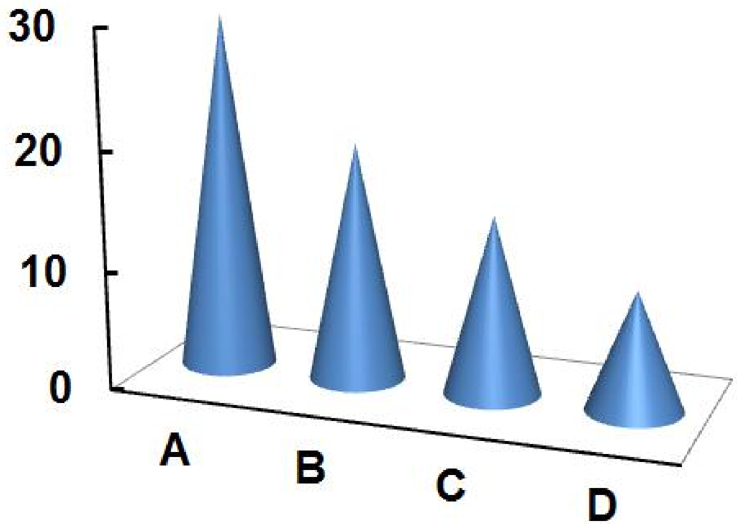


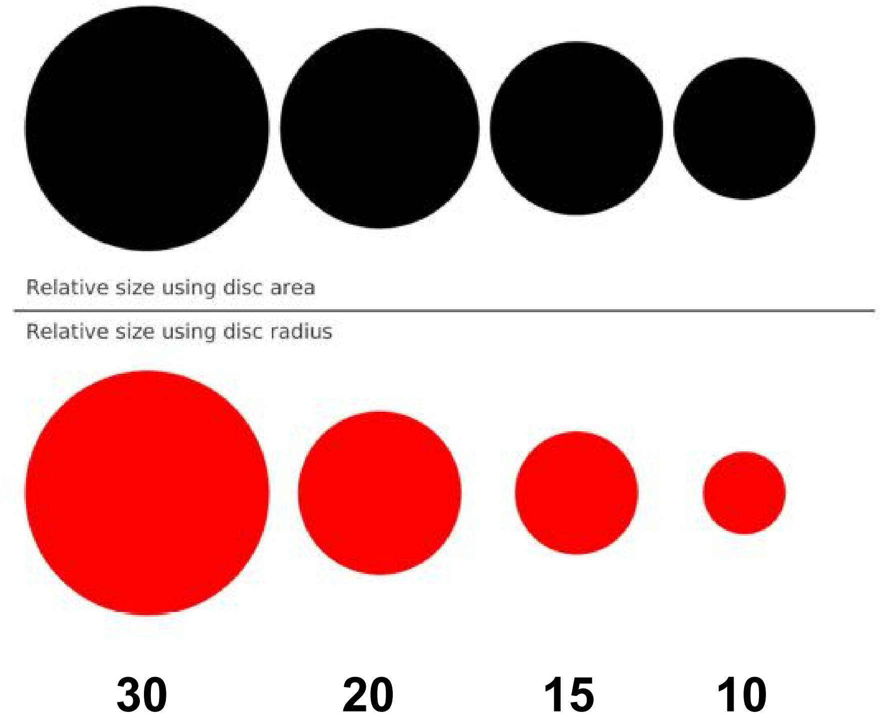
Rolandi et al. (2011), A brief guide to designing effective figures for the scientific paper, *Advanced Materials*

3. Optimize lines and symbols

People are best at detecting differences in length (in one direction)







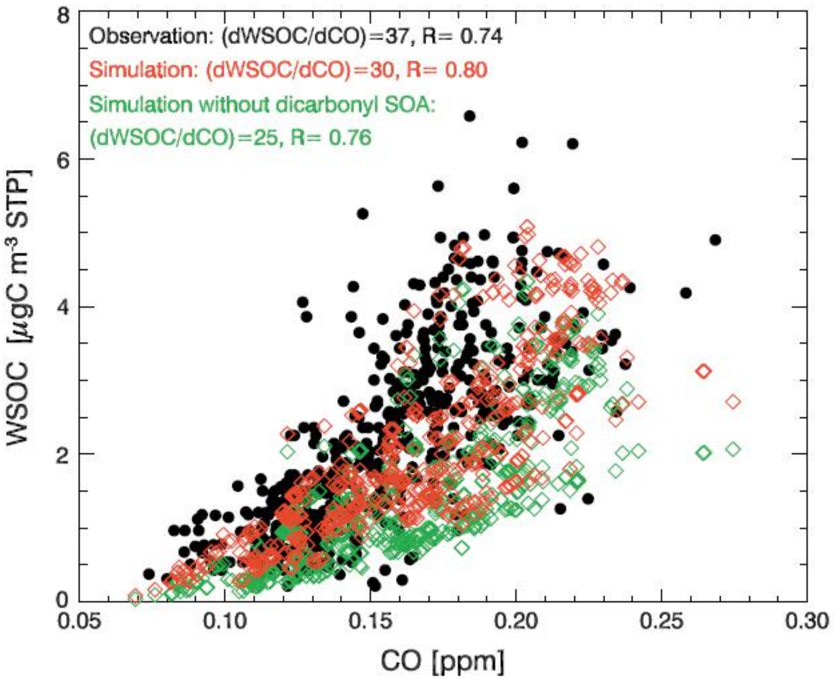
Rougier et al (2014), Ten simple rules for better figures, *PLOS Computational Biology*

3. Optimize lines and symbols

Choose lines & symbols (thickness & sizes) that show data clearly and contrast easily

- **Closed symbols show up better**
- **Open and closed symbols show maximum contrast**
- **Squares and circles look similar**
- **Size: depends on amount of data and precision**

Bad example: symbols obscured

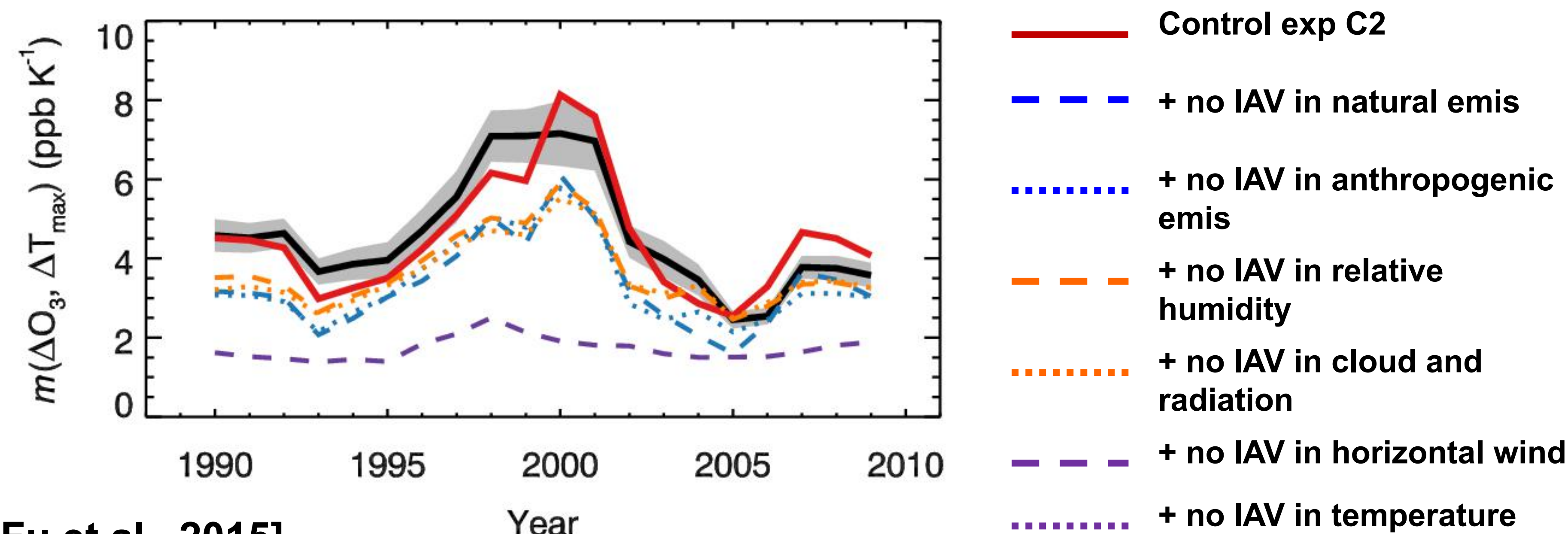


Fu et al. [2009]

3. Optimize lines and symbols

Choose lines that show data clearly and contrast easily

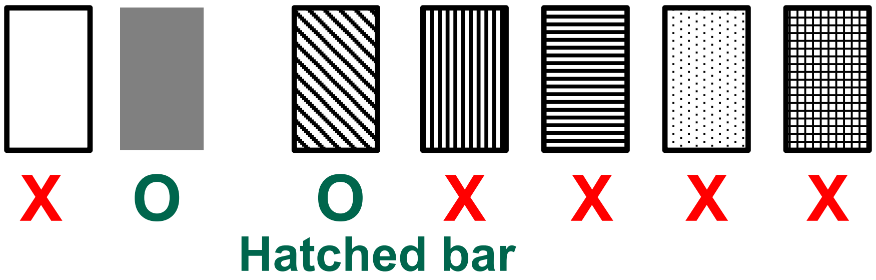
- Keep line styles simple (< 3 styles)
- Make thickness and style differences obvious
- Use thick, solid line to show observations
- Use thinner, dashed lines to show less important results
- Lines should be thick. Do not use hairpin lines anywhere

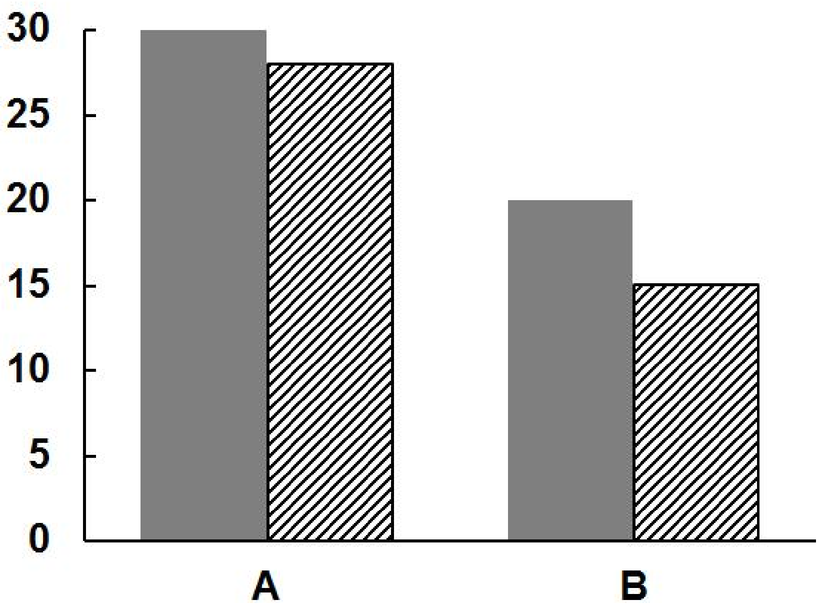


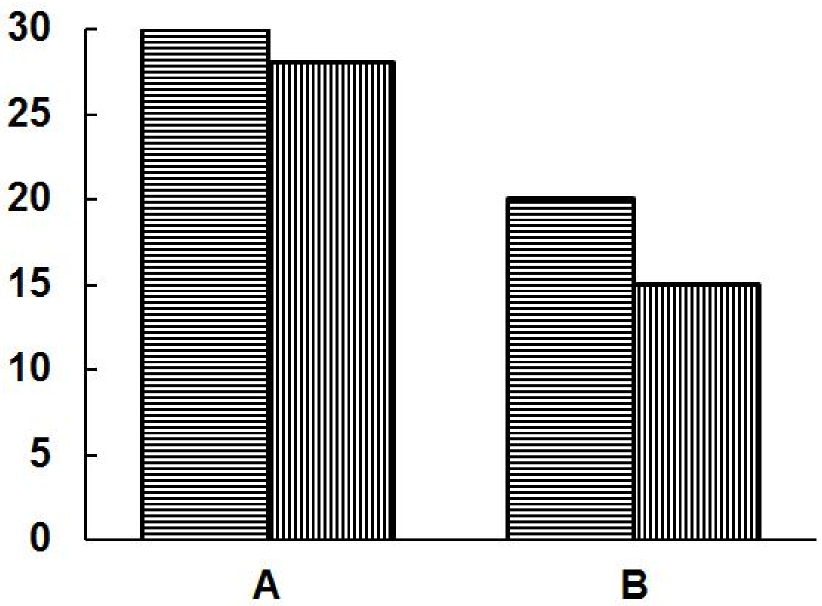
[Fu et al., 2015]

3. Optimize lines and symbols

- Use <3, simple texture patterns



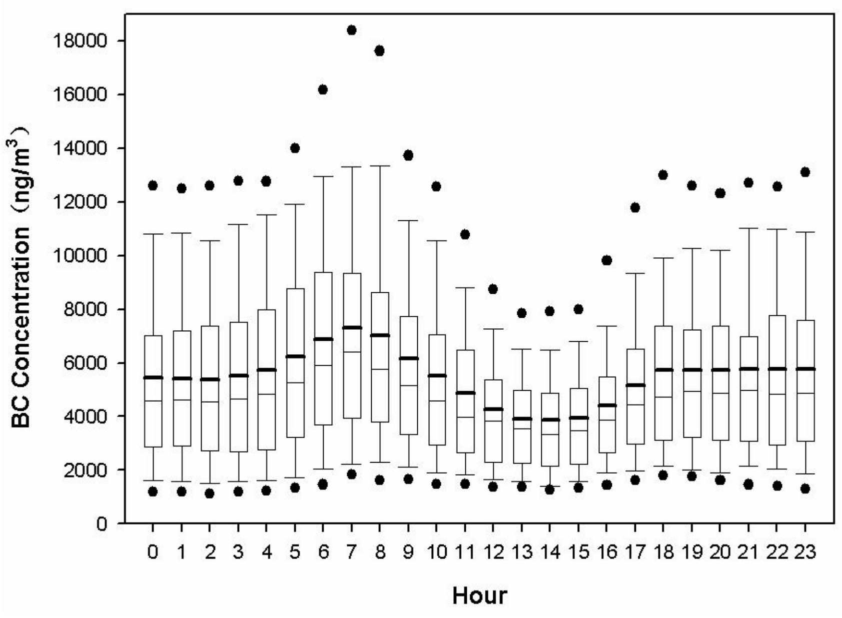


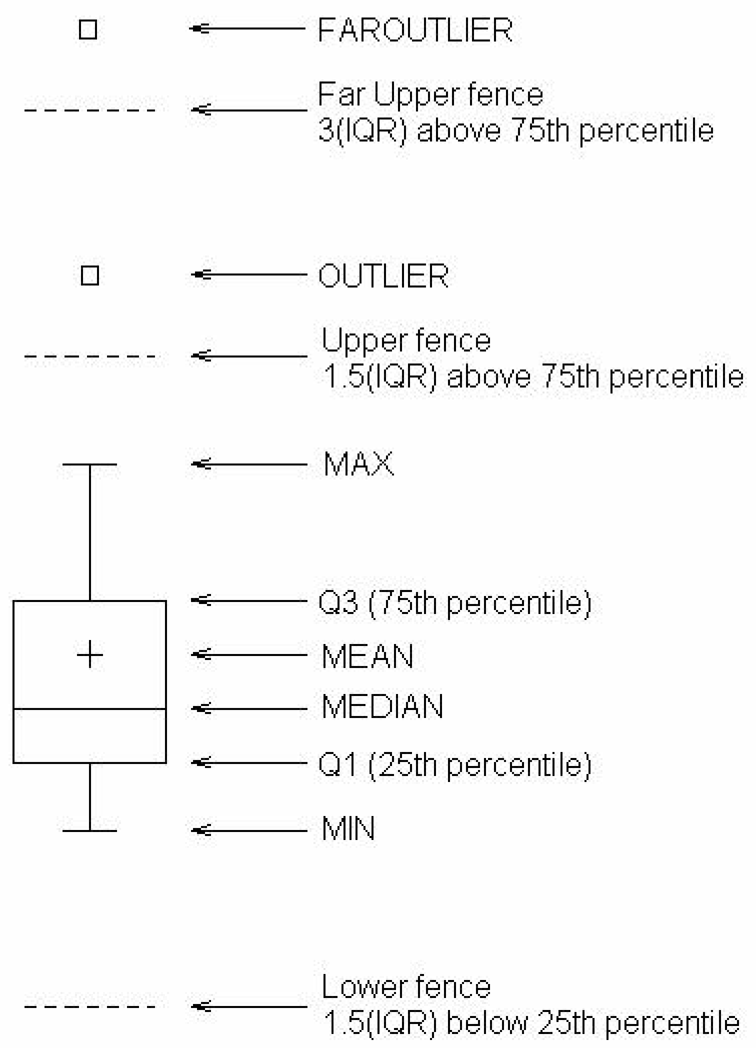


3. Optimize lines and symbols

Presenting uncertainties:

1. Box plot: mean, median, 25th & 75th percentiles, outliers

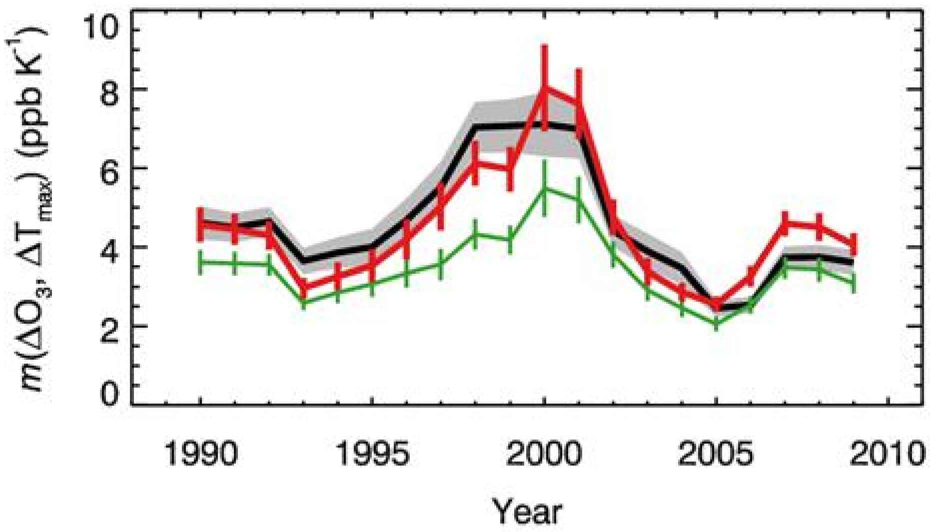




3. Optimize lines and symbols

Presenting uncertainties:

2. Error bars: standard deviation, 95% confidence interval, range



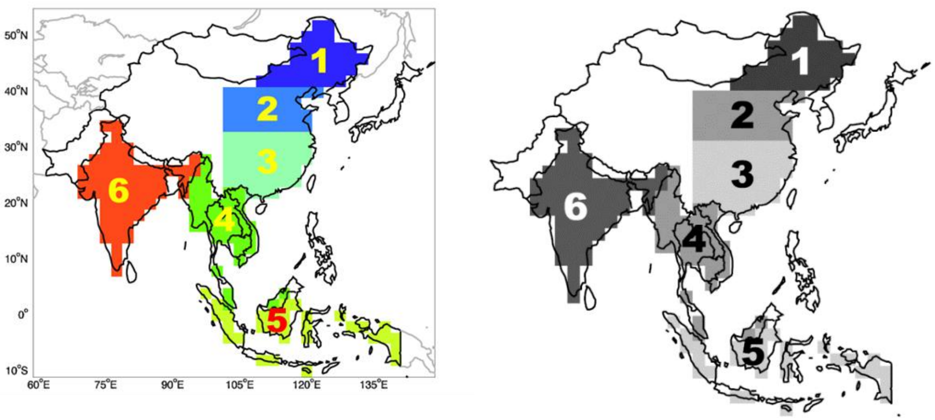
Whiskers/shading: standard errors of the slopes

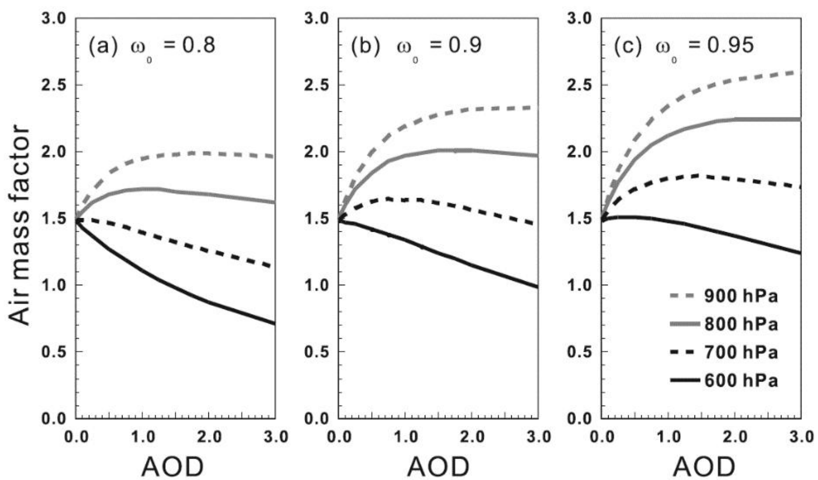
3. Ensemble results: permutation uncertainty

Fu et al. [2015]

4. Optimize colors and color scales

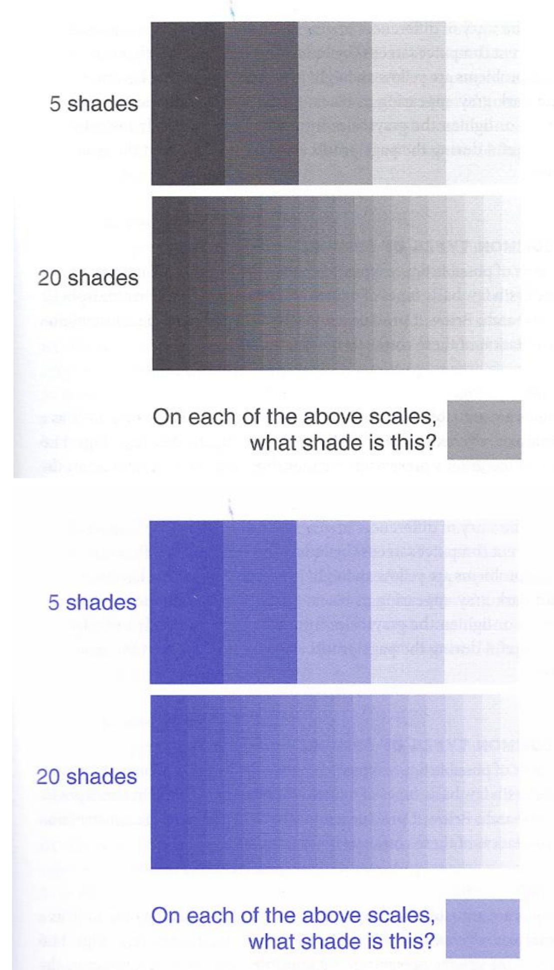
- Use colors only if necessary
- Instead use B/W or monotone shades to reduce cost and to maintain figure quality for B/W printing





Fu et al. [2007]

- Do not use more than 5 shades for quantification

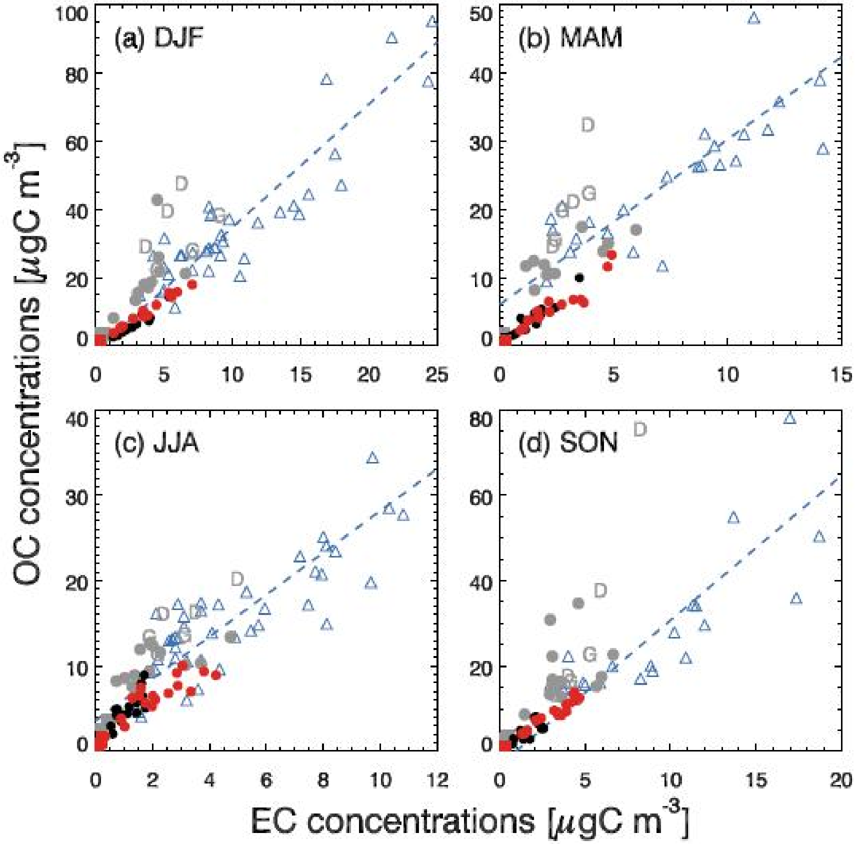


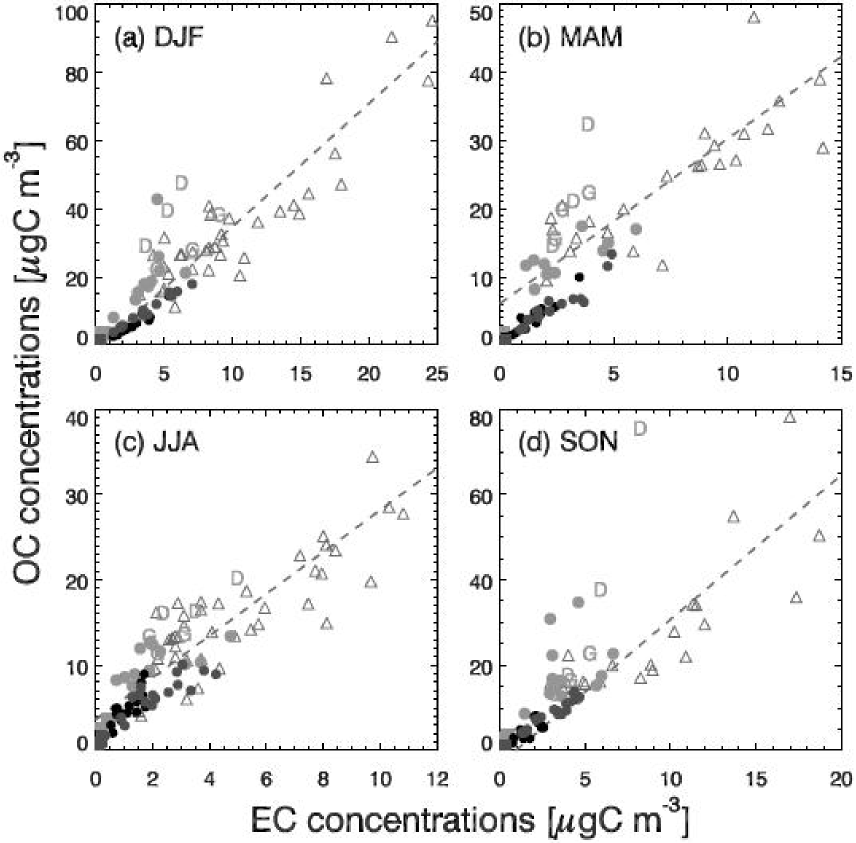
Schultz (2009), *Eloquent Science*

4. Optimize colors and color scales

Best practice: use color and patterns that reproduce well in B/W, although this is not always possible with > 3 colors

Bad example: here, red and black can be distinguished, but blue and grey get mixed-up





Fu et al. [2012]

4. Optimize colors and color scales

Use unambiguous, saturated colors for lines and symbols

4. Optimize colors and color scales

Use sufficient contrast, use color if necessary

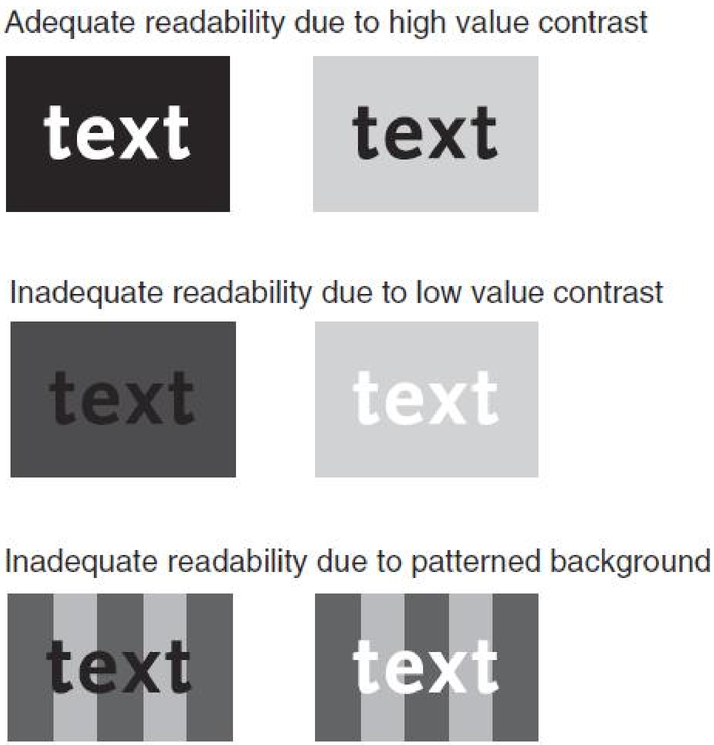
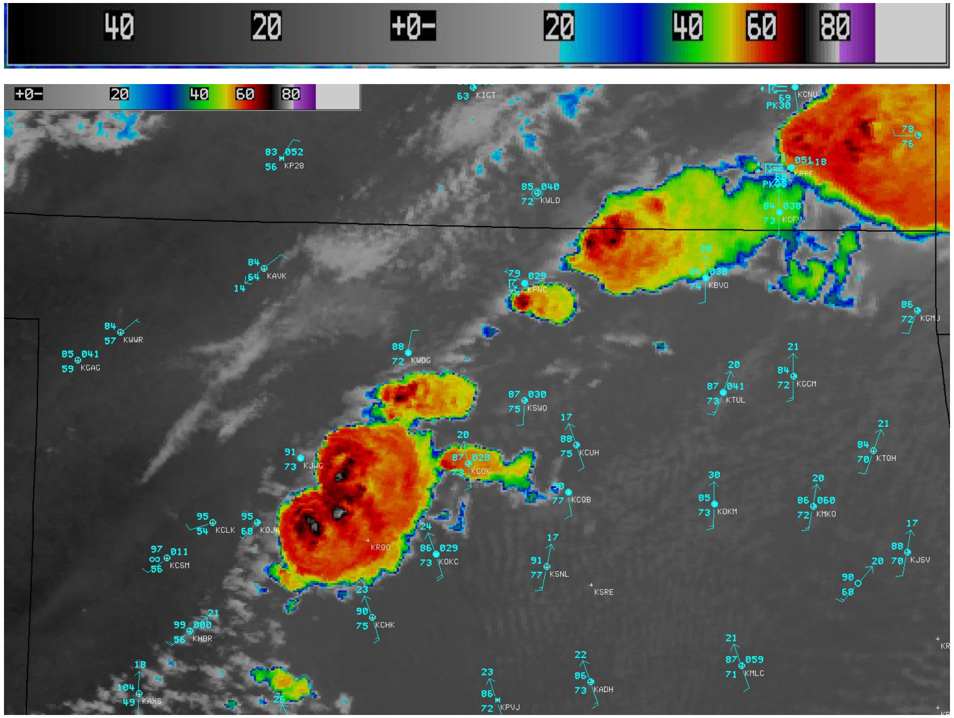


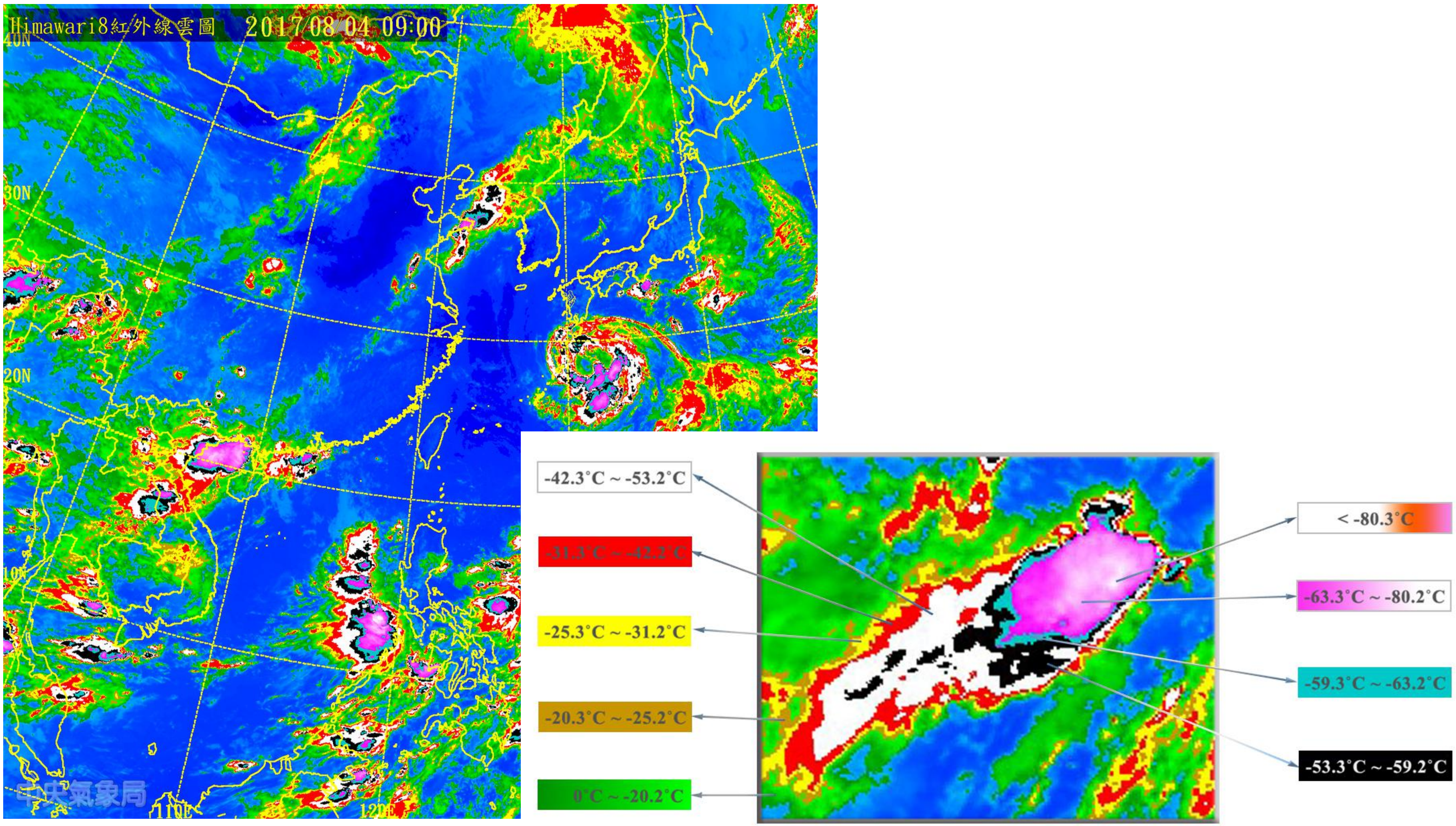
Figure 5. Strive for at least 70% contrast between text and background. Avoid placing text over a patterned background; instead, text should be placed nearby or adjacent to the image or in the legend.



Rolandi et al. (2011), A brief guide to designing effective figures for the scientific paper, *Advanced Materials*

4. Optimize colors and color scales

Abrupt color jumps qualitatively emphasize thresholds and hotspots



4. Optimize colors and color scales

Bad color scale: abrupt color jumps make it hard to detect differences quantitatively

I made this plot and published it ☹

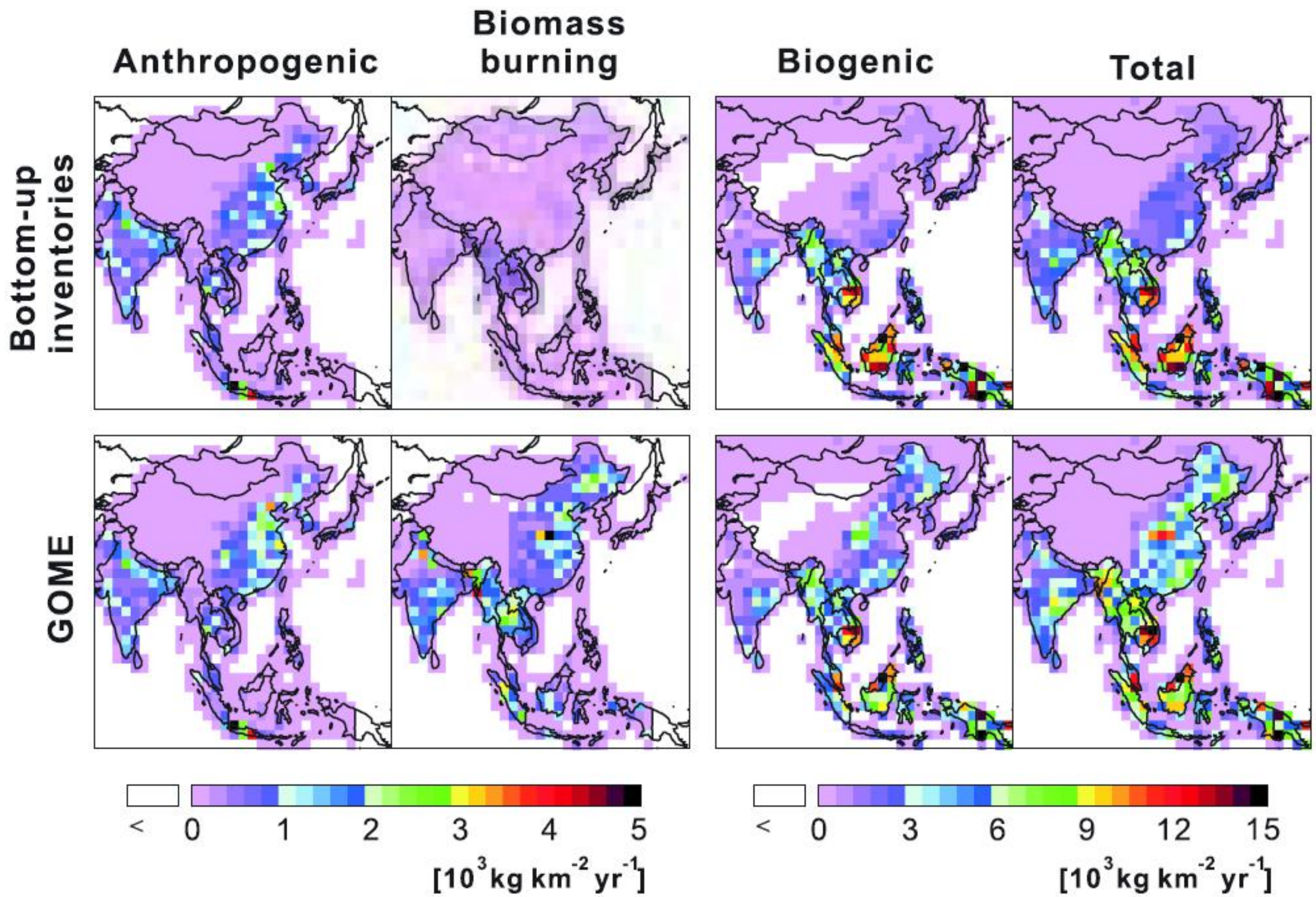
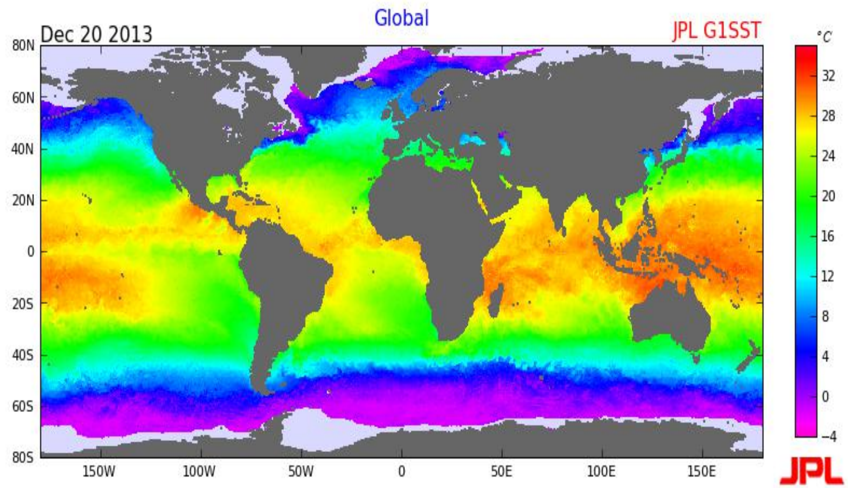


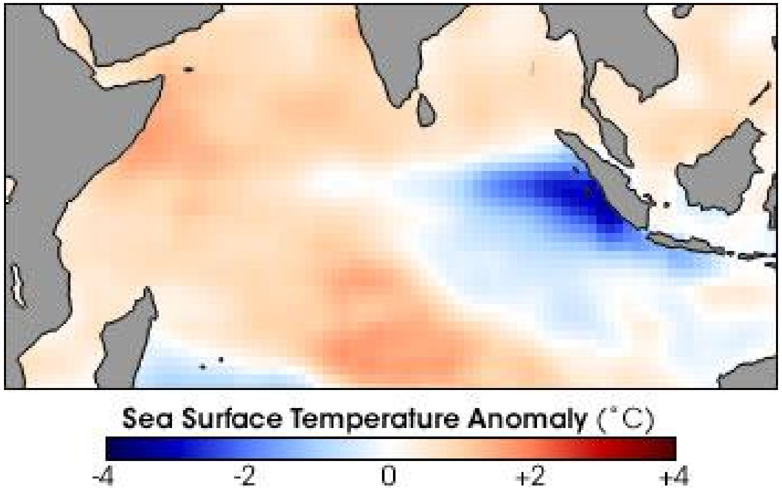
Figure 3. Annual mean reactive NMVOC emissions from east and south Asia. (top) Bottom-up inventories of *Streets et al.* [2003a] (anthropogenic, biomass burning) and *Guenther et al.* [2006] (biogenic). (bottom) Emissions inferred from GOME HCHO on the basis of the regression analysis from this study. Color scale at the left indicates anthropogenic and biomass burning, and the color scale on the right indicates biogenic and total sources.

Fu et al. [2007]

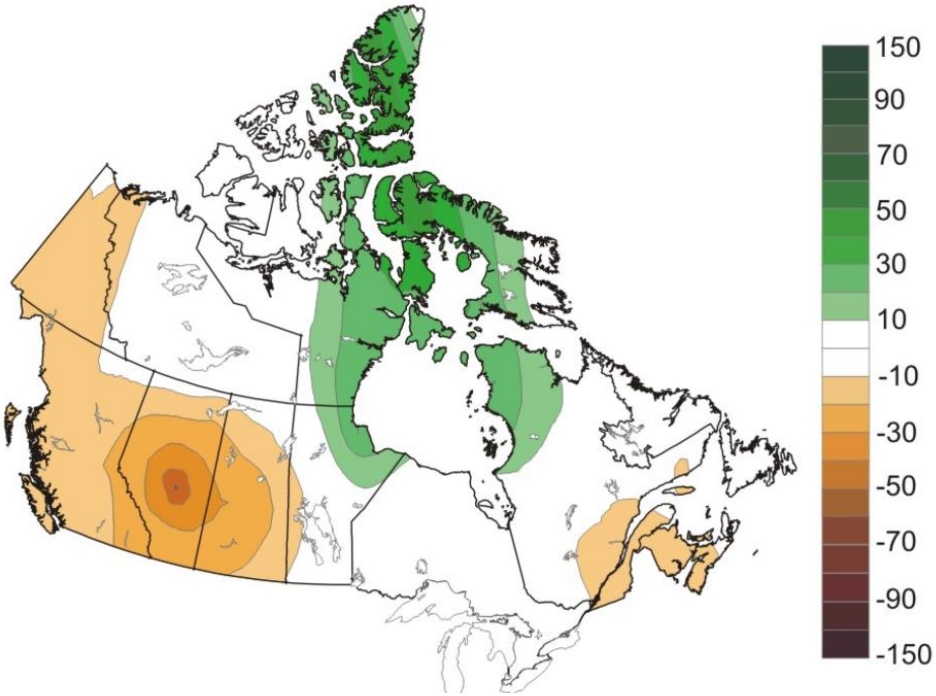
4. Optimize colors and color scales

Try to use colors that make physical sense



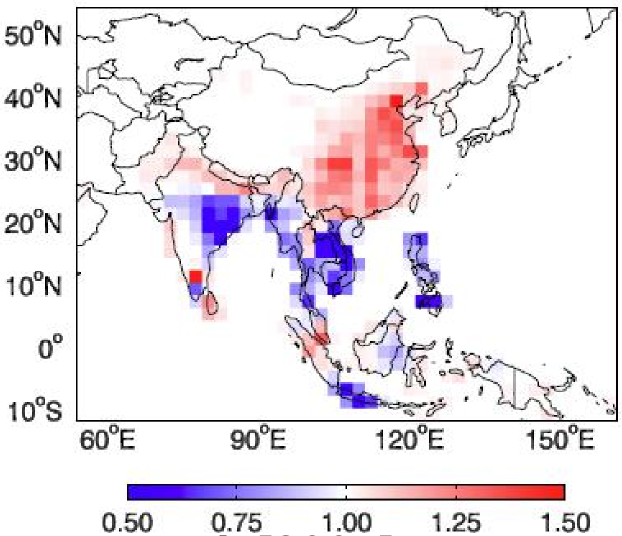


Precipitation anomaly over Canada

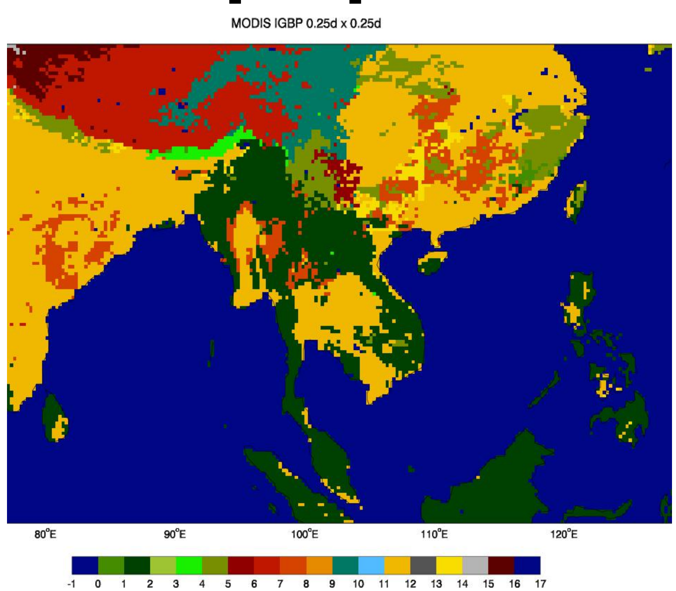


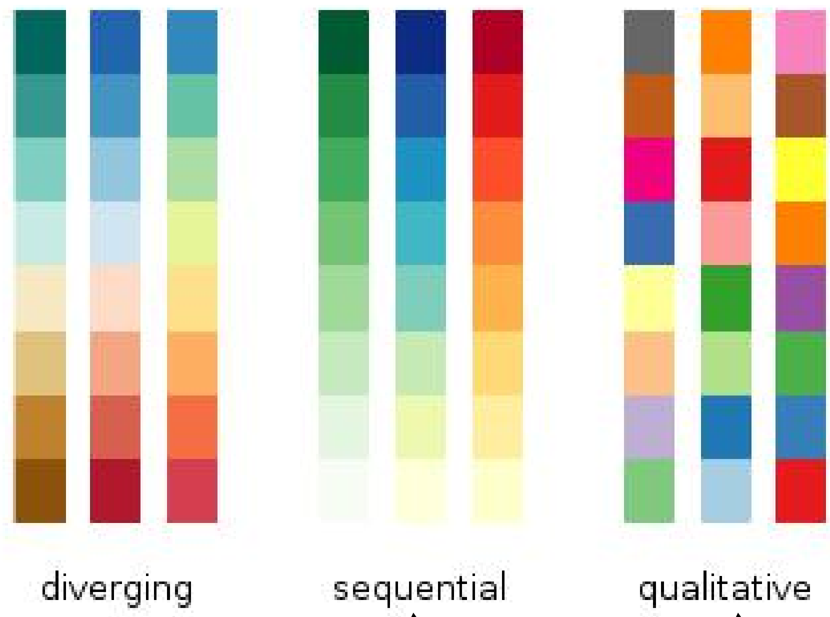
4. Optimize colors and color scales

CO emission scale factor



Kopacz et al. [2007]

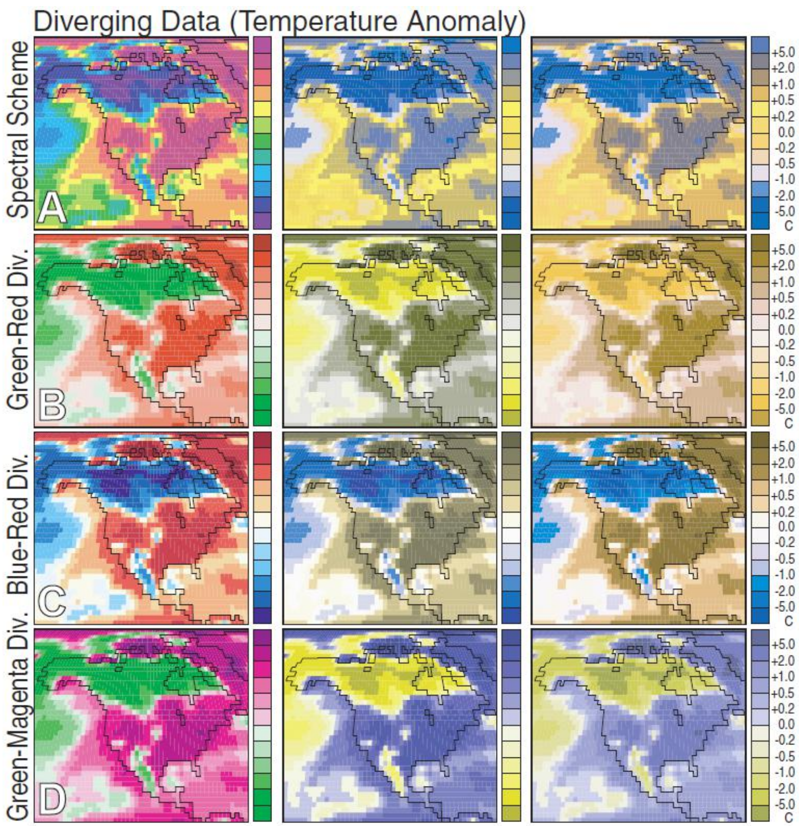


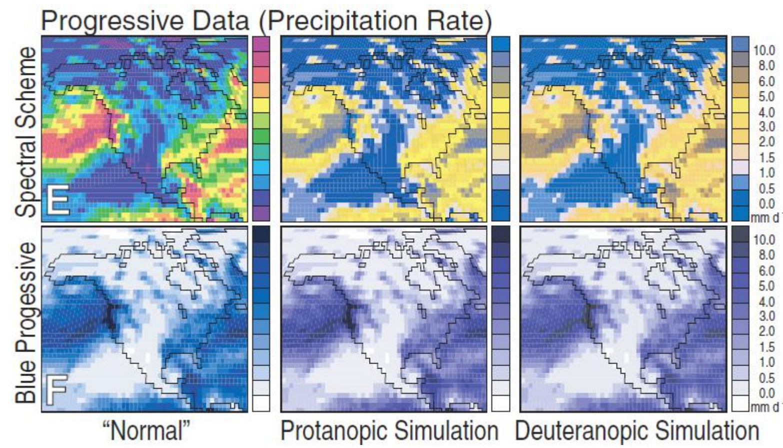


numerical variables

categorical variables

4. Optimize colors and color scales



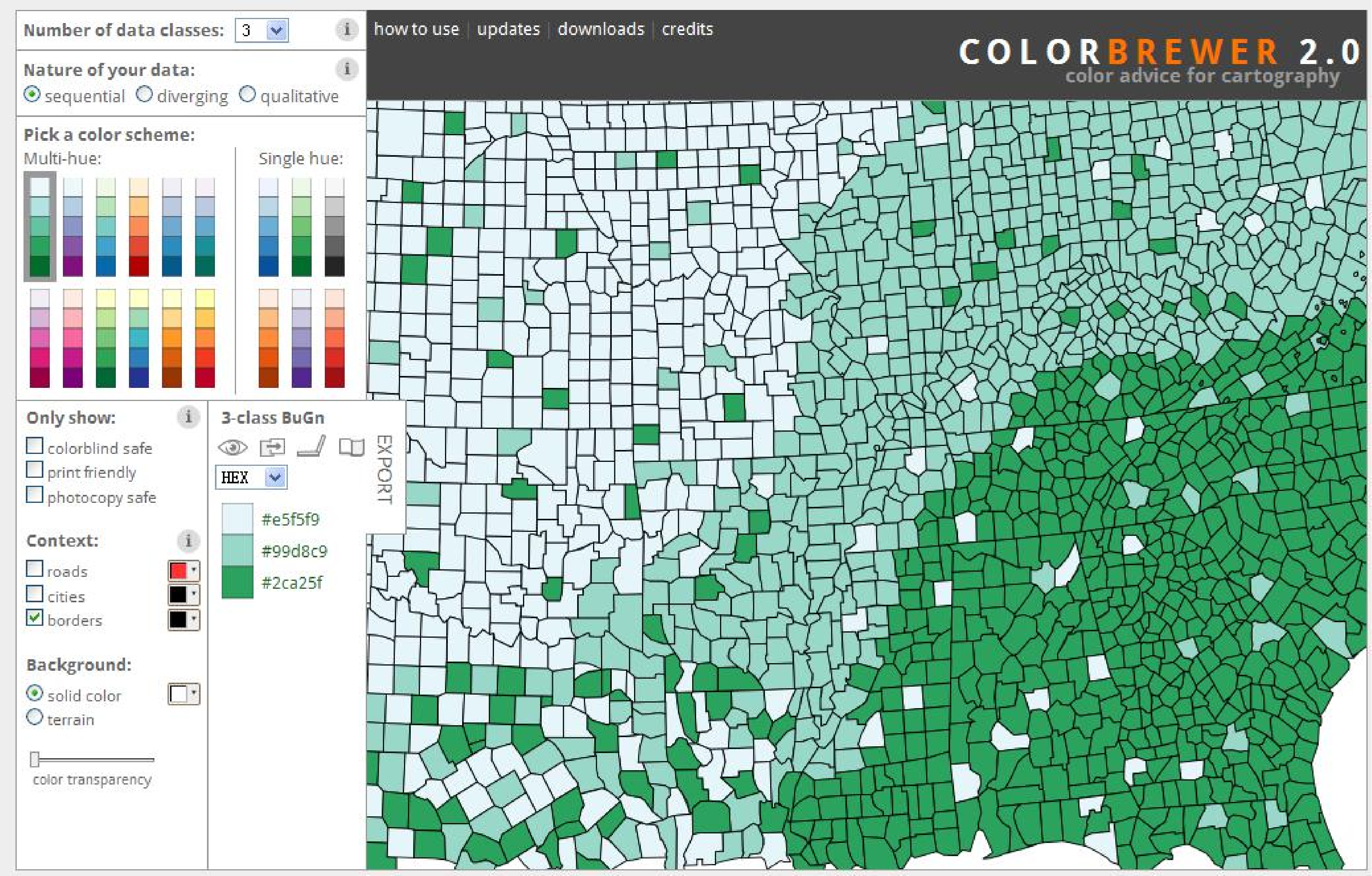


Red-green colorblind **Green colorblind**

Red-green colorblind **Green colorblind**

Light and Bartlein (2004), The End of the Rainbow? Color Schemes for Improved Data Graphics, *EOS*

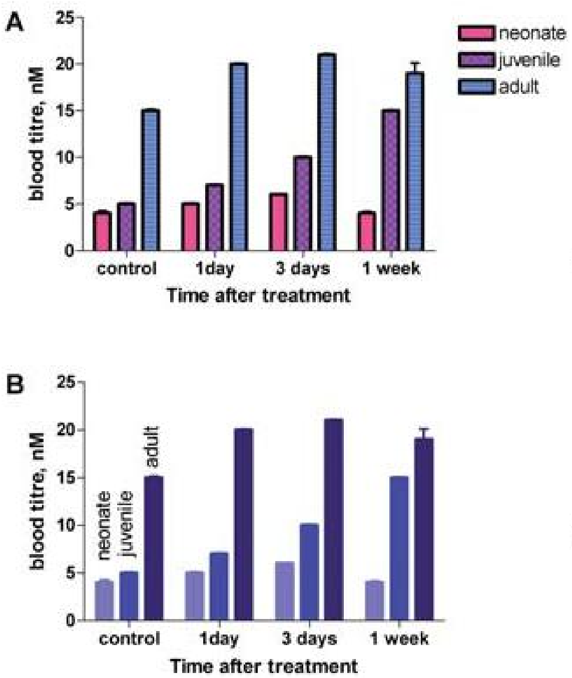
4. Optimize colors and color scales

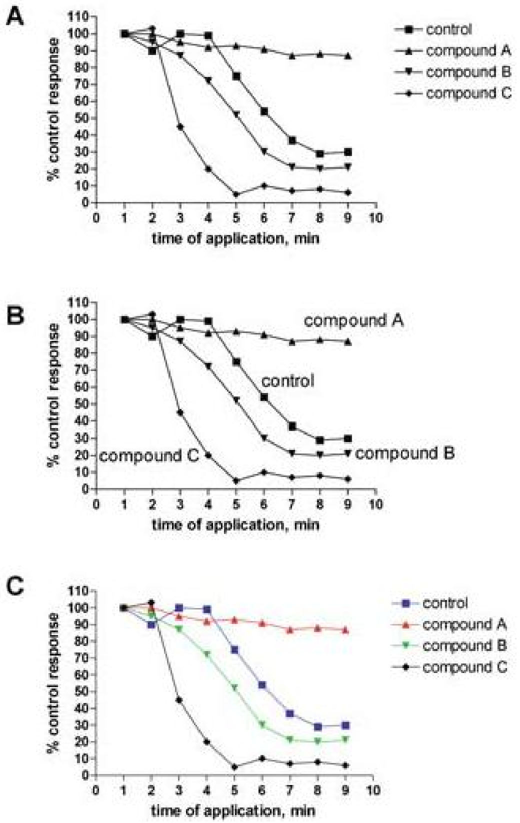


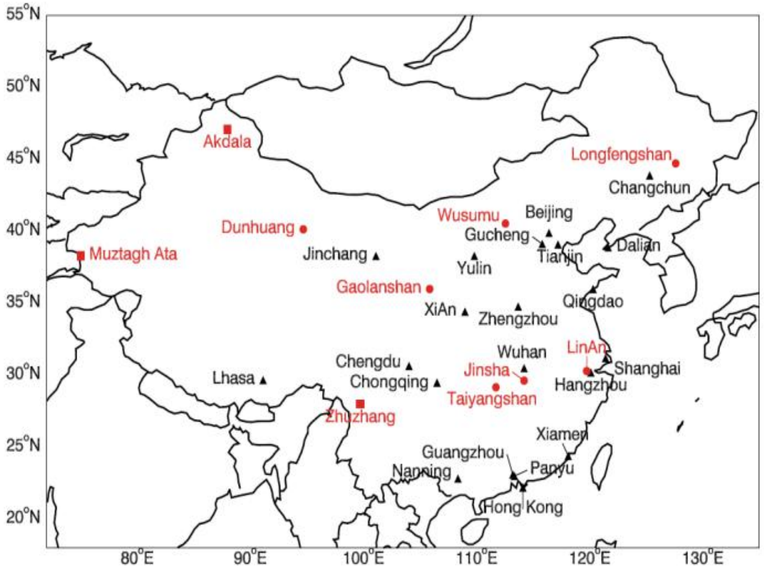
colorbrewer2.org
vischeck.com

5. Optimize labels and legends

Try to label the figure, not the legend, unless doing so clutters the figure too much



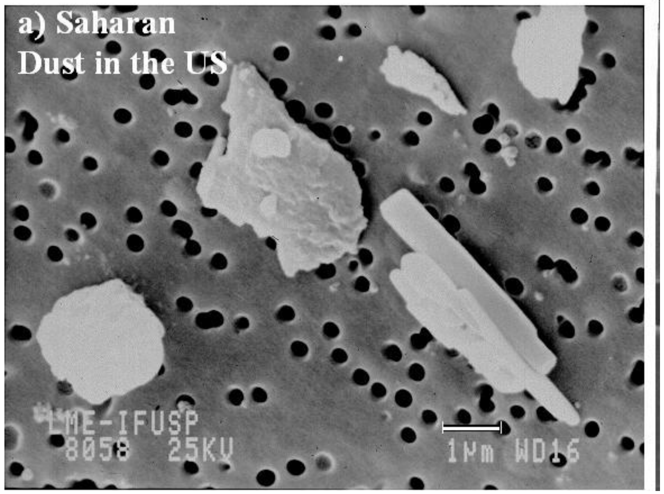




Buckingham (2008), Effective scientific illustrations, *Lab Times*

5. Optimize labels and legends

- All elements and axes should be clearly labeled
- Avoid changing orientation of labels
- Avoid acronyms and abbreviations in labels and legends, unless necessary and defined in text
- Numeric labels should show appropriate significant figures
- Units should be in MKS and written as [10^3 μg cm^{-2} s^{-1}]:
  NOT [1000 mug/grid/s]
  NOT [1000 mg/cm^2 • s]
  NOT [thousand ug/s cm^2]
- 36.7 ^oC, 80 ^oF, BUT 273.15 K
- Do not draw box around the legend
- Scales should be hard-wired inside photo, next to features



6. Use consistent fonts and font sizes

Use Arial (Helvetica) for English, 黑体 for Chinese, symbol for Greek letters

Serif font (Times, 24 pt)　宋体
Sans serif font (Arial, 24 pt)　黑体

Original resolution

Serif font (Times, 24 pt)　宋体
Sans serif font (Arial, 24 pt)　黑体

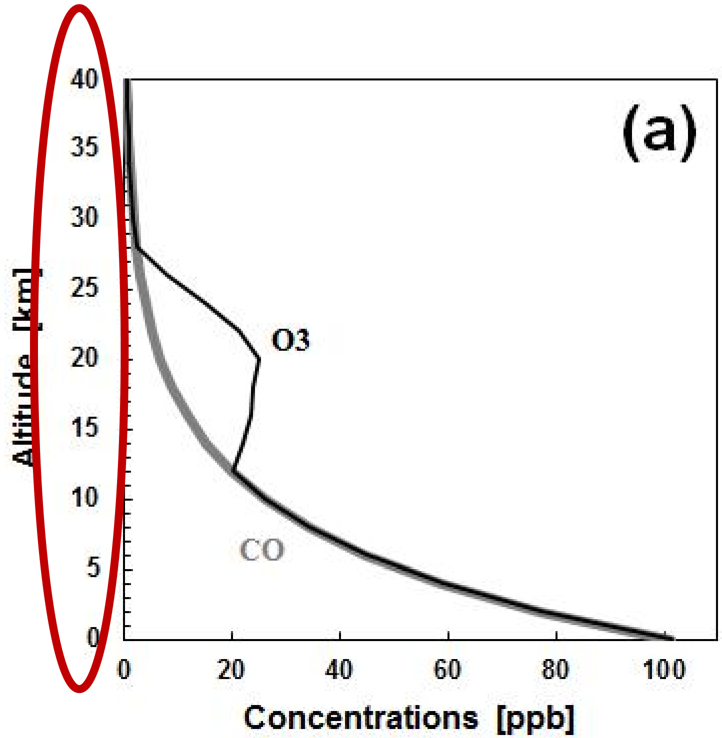
Reduced resolution

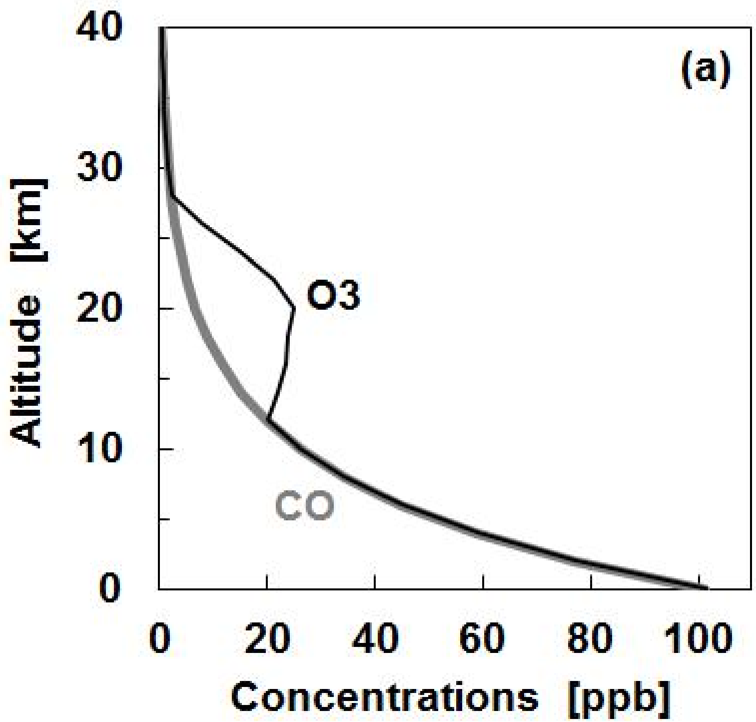
Serif font (Times, 12 pt)　宋体
Sans serif font (Arial, 12 pt)　黑体

Reduced font size

Special **fonts** *may* vοτ *show* **up the way** you intended!

6. Use consistent fonts and font sizes

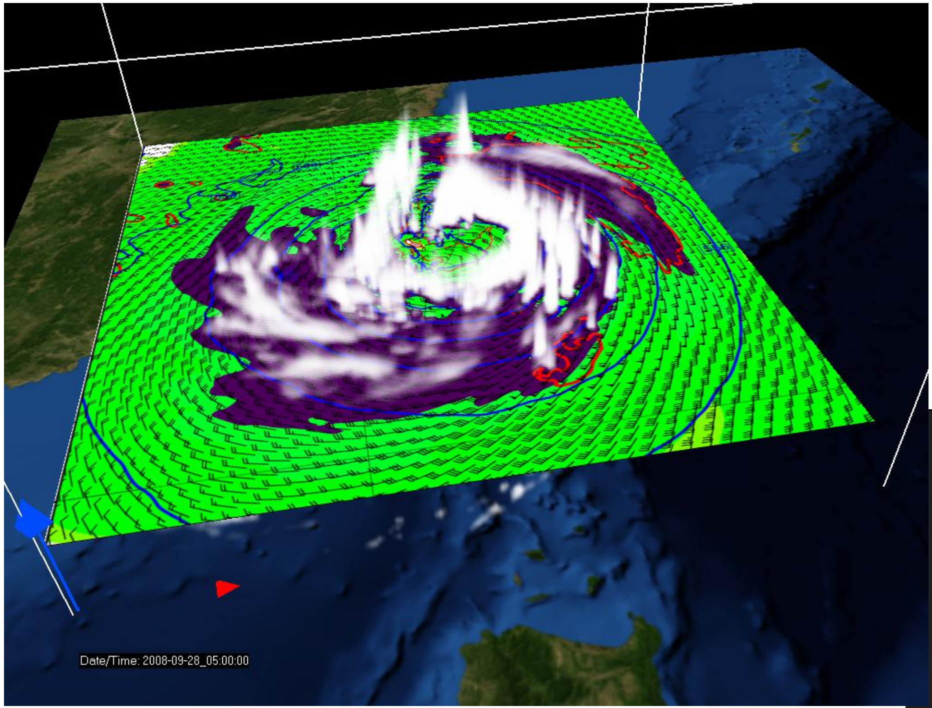


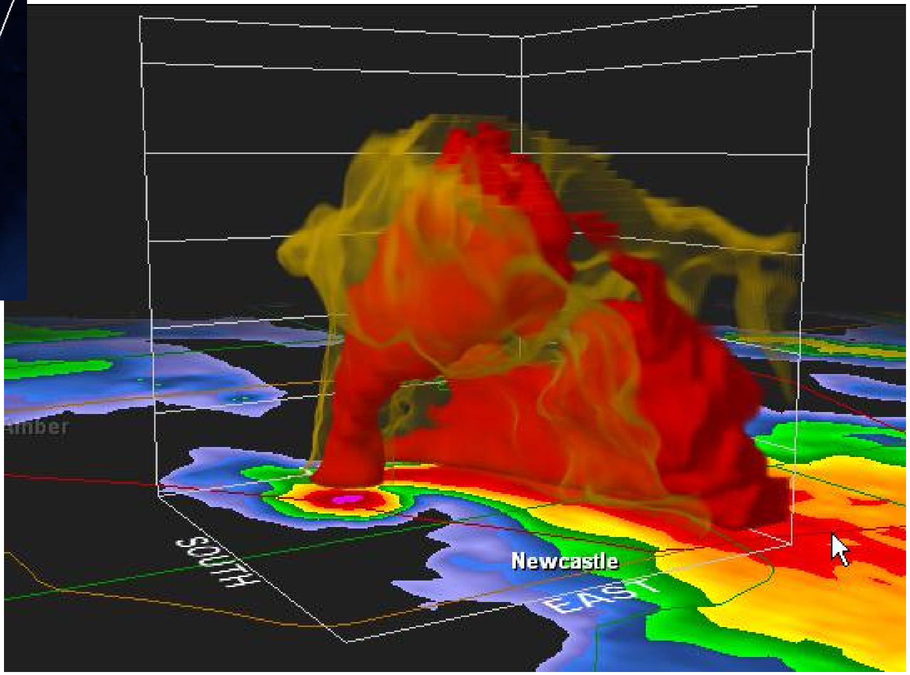


Number labels often tend to be too small

7. Use 3-D effects and animations with caution

3-D effects usually only useful for presenting qualitative information

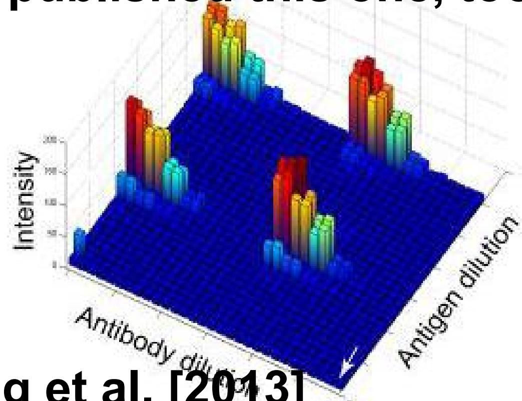




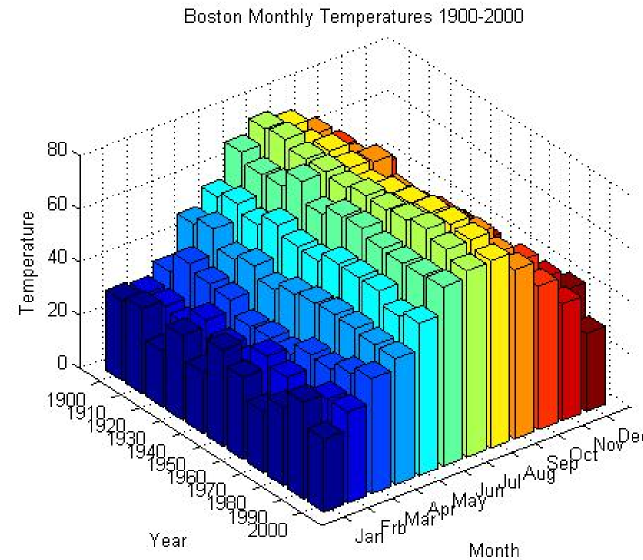
7. Use 3-D effects and animations with caution

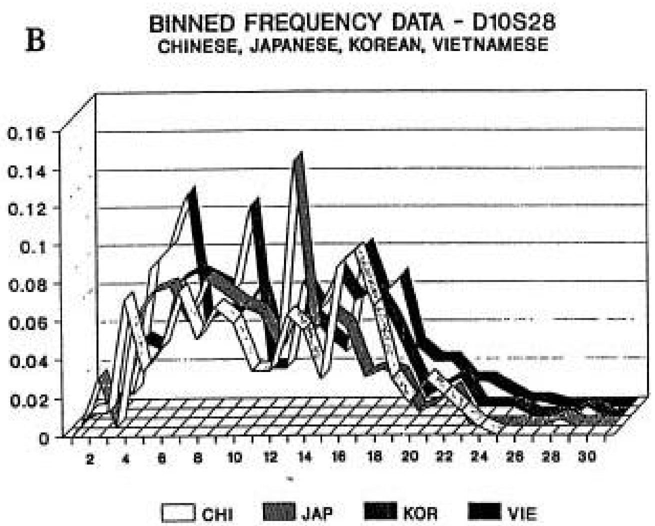
More often than not, 3-D effects are useless or even confusing

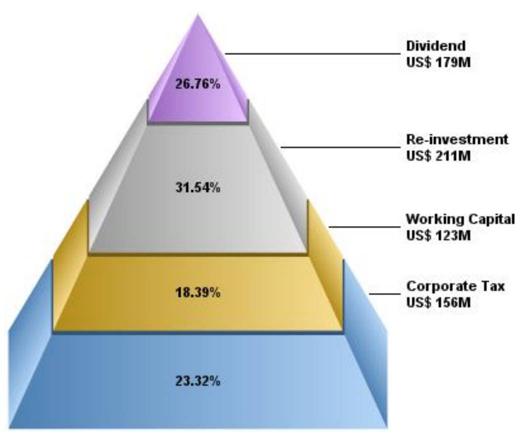
I published this one, too ☹

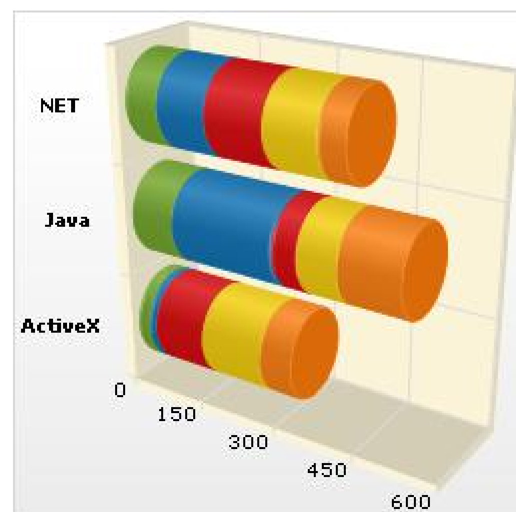


Zhang et al. [2013]

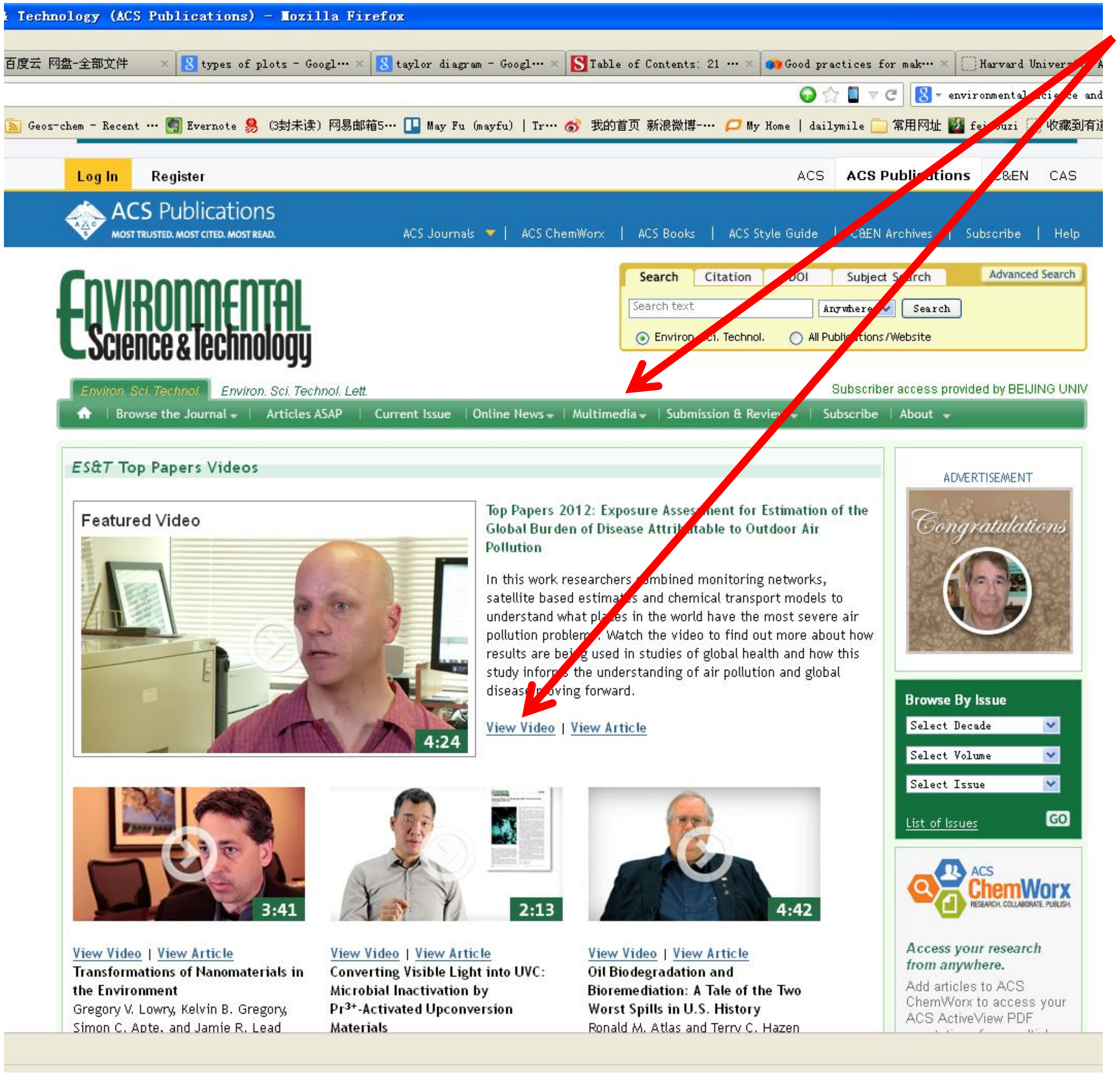








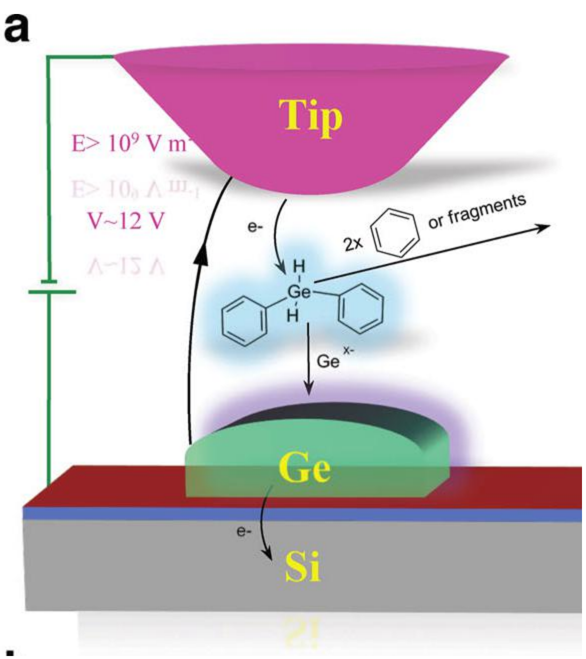
7. Use 3-D effects and animations with caution

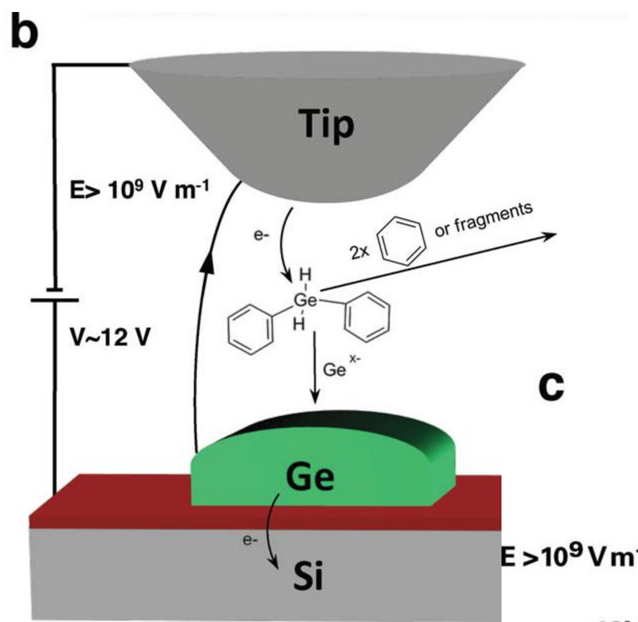


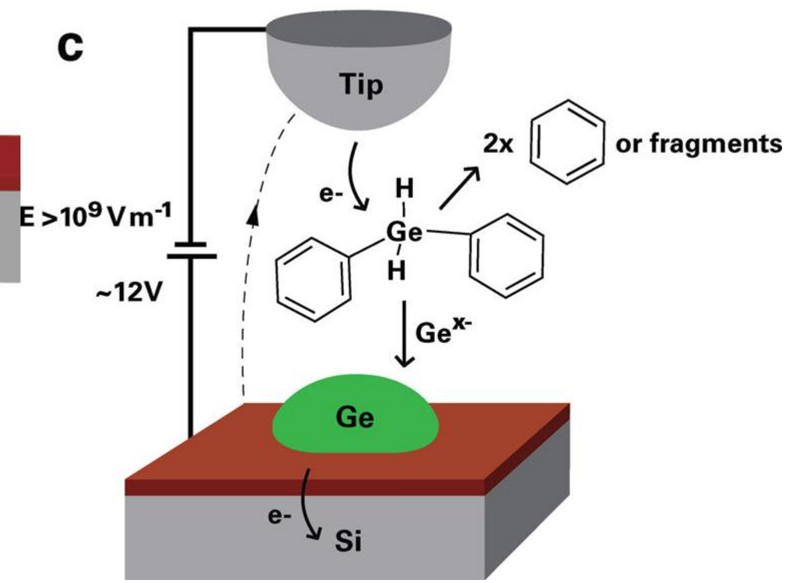
- **Several journals now allow multimedia in online supplementary materials.**
- **Animations and movies can be welcome breaks in presentations, if and only if they play correctly.**

8. Focus on key message and simplify

Maximize data-to-ink ratio!



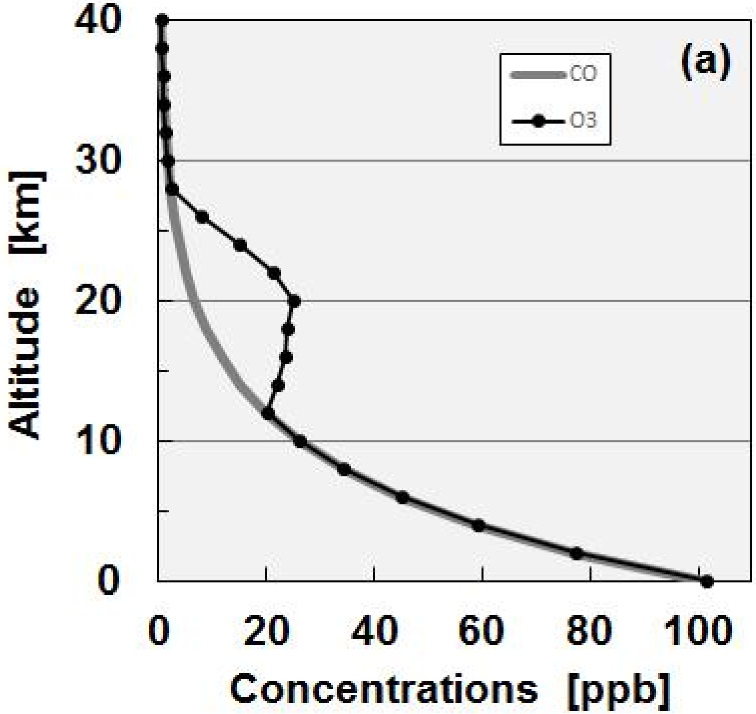


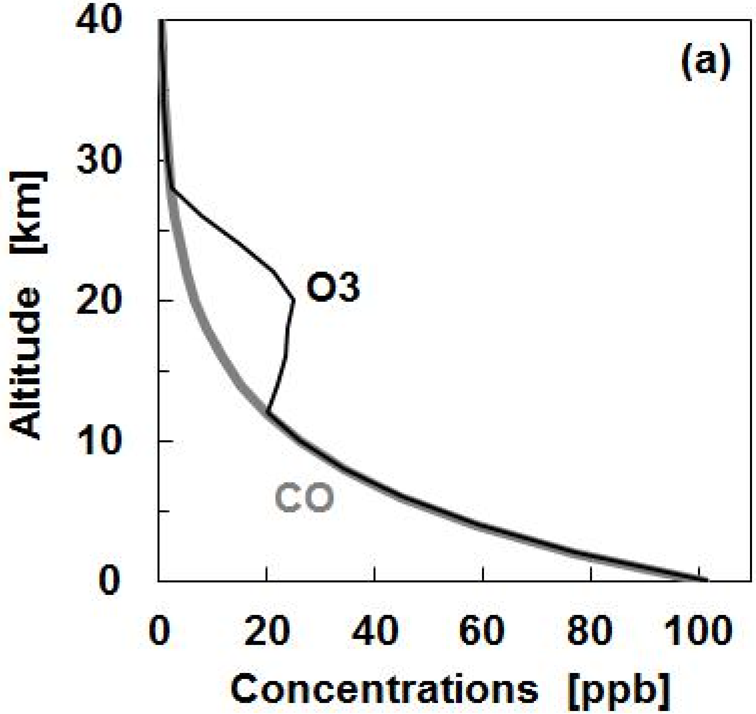


Rolandi et al. (2011), A brief guide to designing effective figures for the scientific paper, *Advanced Materials*

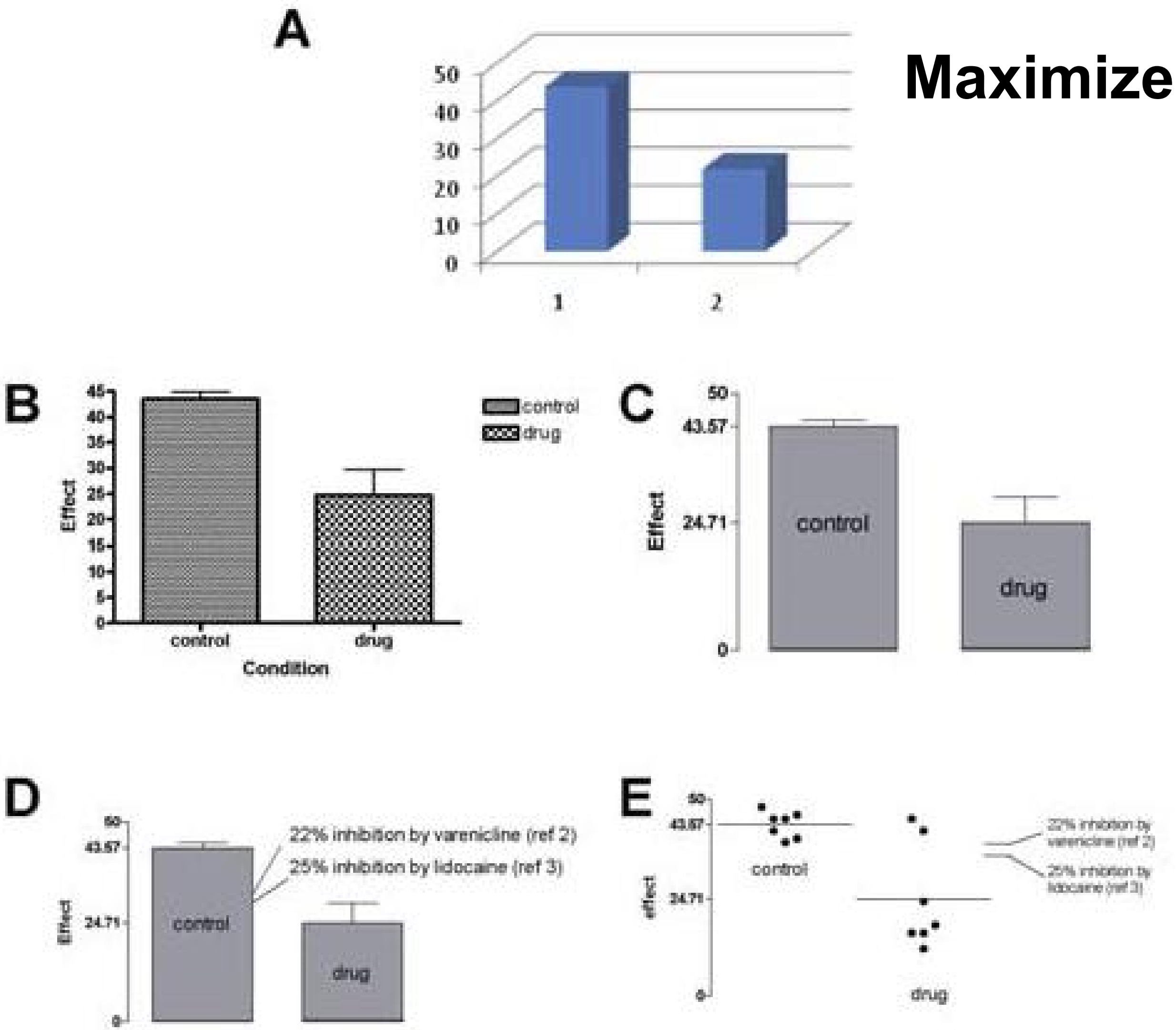
8. Focus on key message and simplify

Maximize data-to-ink ratio!





8. Focus on key message and simplify



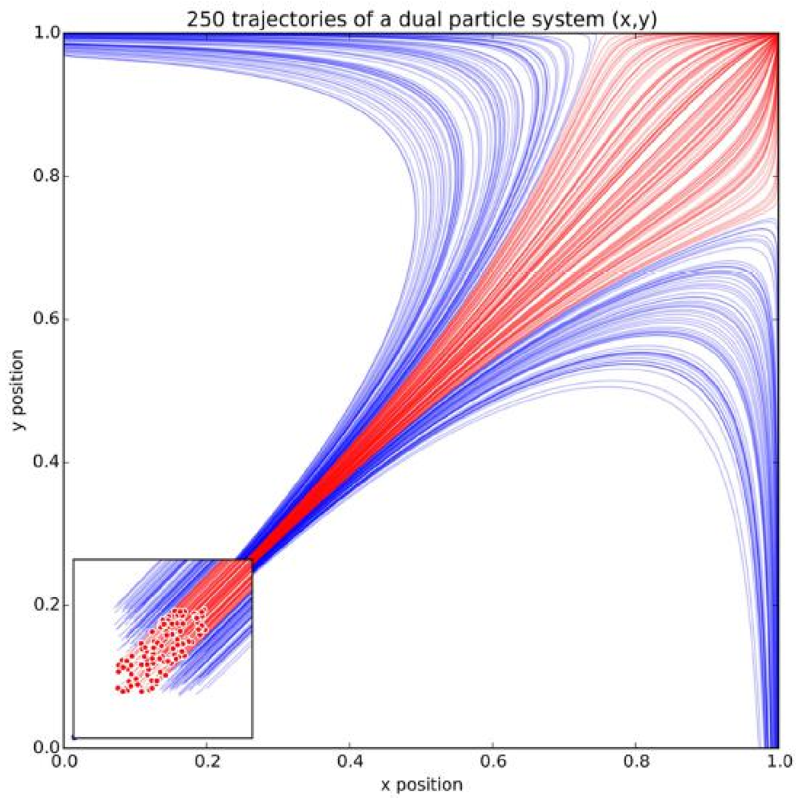
Maximize data-to-ink ratio!

Buckingham (2008), Effective scientific illustrations, *Lab Times*

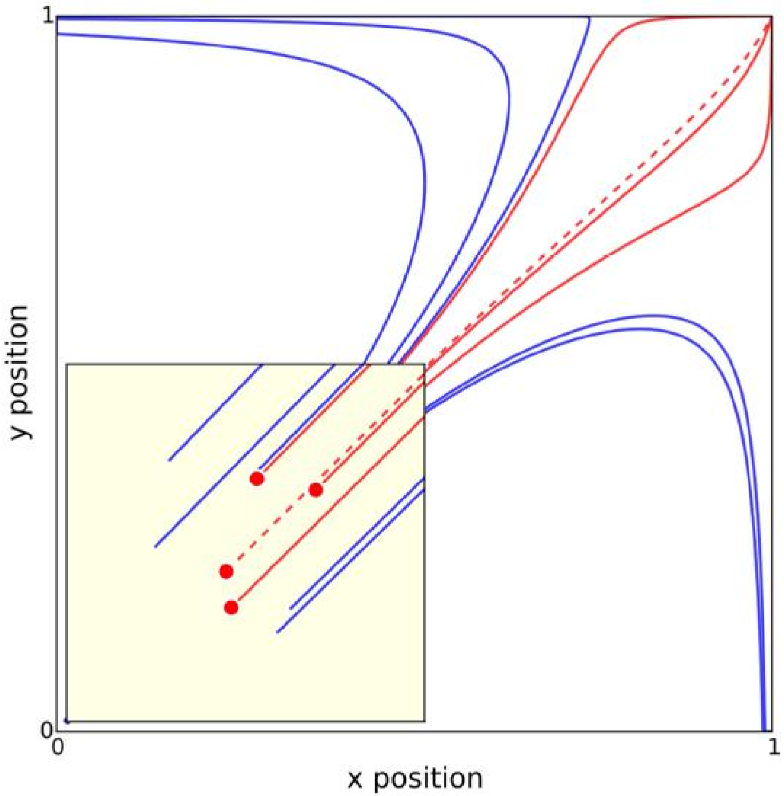
8. Focus on key message and simplify

Adapt to the support medium

Journal



Oral presentation



Rougier et al (2014), Ten simple rules for better figures, *PLOS Computational Biology*

Adapt to the support medium

Table 4. Global Sources of Secondary Organic Aerosol in GEOS-Chem[a]

Mechanism	Reversible	Reversible	Reversible	Irreversible	
Precursors	Terpenes, Terpenoid Alcohols, and Sesquiterpenes	Isoprene	Aromatics[b]	Glyoxal and Methylglyoxal	Total
Source, Tg C a^{-1}	9.0	6.7	1.7	11[c]	29
Above 1.5 km	3.1	2.8	0.95	4.4	
Below 1.5 km	5.9	3.9	0.72	6.2	
Annual mean burden, Tg C	0.19	0.21	0.04	0.17[d]	0.62
Above 1.5 km	0.14	0.18	0.03	0.12	
Below 1.5 km	0.05	0.03	0.01	0.05	

[a]Including reversible formation from condensation of semivolatile products of VOC oxidation, and irreversible formation from oxidation and/or oligomerization of glyoxal and methylglyoxal in aqueous aerosols and clouds. Global production rates of each aerosol type formed by the reversible pathway are derived from balance with deposition. The organic to carbon mass ratio for all reversible-partitioning SOA is assumed to be 2.1, following *Henze et al.* [2008].

[b]From *Henze et al.* [2008].

[c]Includes 2.6 Tg C a^{-1} from glyoxal and 8 Tg C a^{-1} from methylglyoxal (Table 2).

[d]Includes 0.04 Tg C from glyoxal and 0.13 Tg C from methylglyoxal (Table 2).

Fu et al. [2008]

9. Write full captions from early on

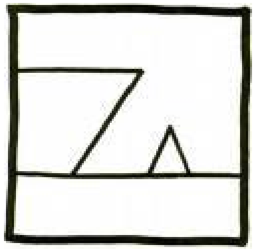


Figure 1. Ship arriving too late to save a drowning witch (Zappa, 1982).

9. Write full captions from early on

Paulot et al. (2009), *Science*

Fig. 4. Simulated daily distribution of IEPOX in the planetary boundary layer during the Northern Hemisphere summer (**A**) and winter (**B**). IEPOX seasonal cycle mirrors the isoprene emissions. The mixing ratio of IEPOX is higher in the tropics than in other isoprene production regions in the northern mid-latitudes (e.g., the southeast United States). This reflects the reduction in the yield of IEPOX from isoprene due to anthropogenic emissions of NO.

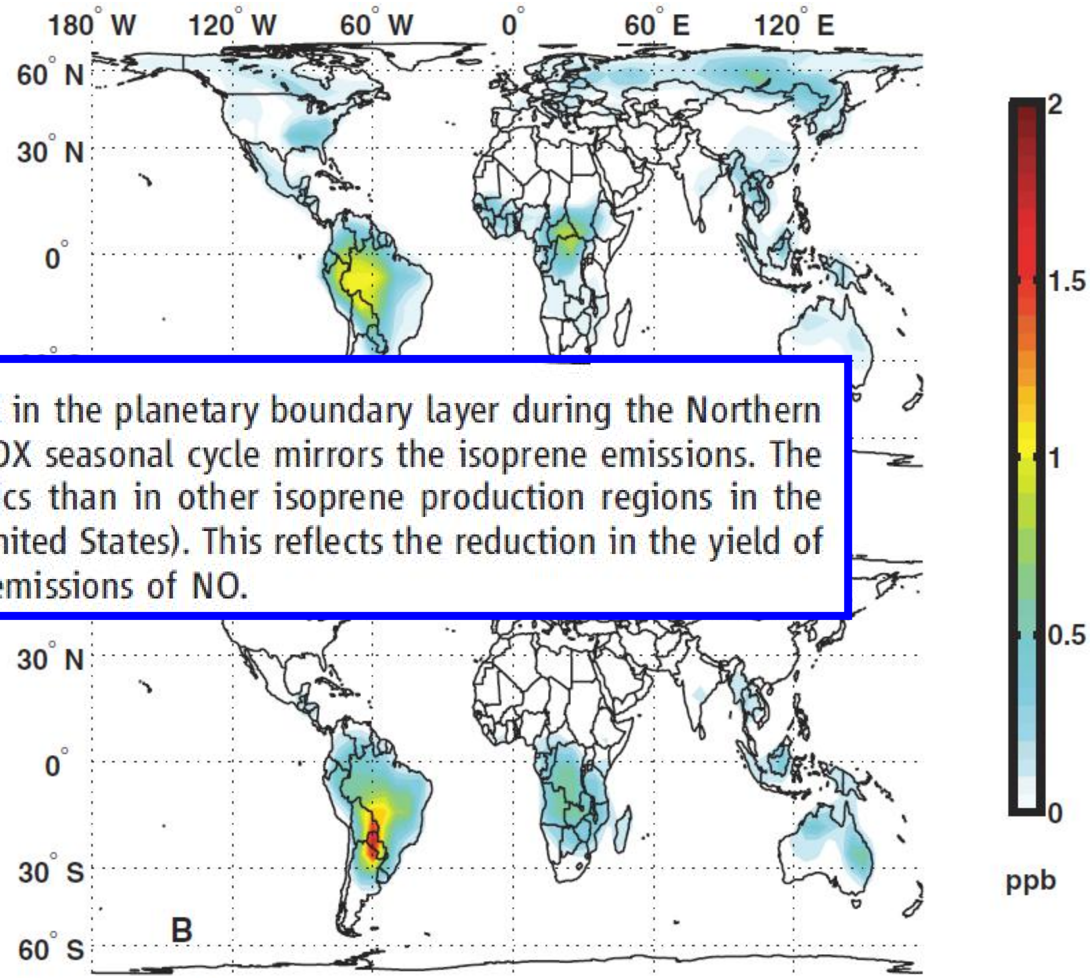
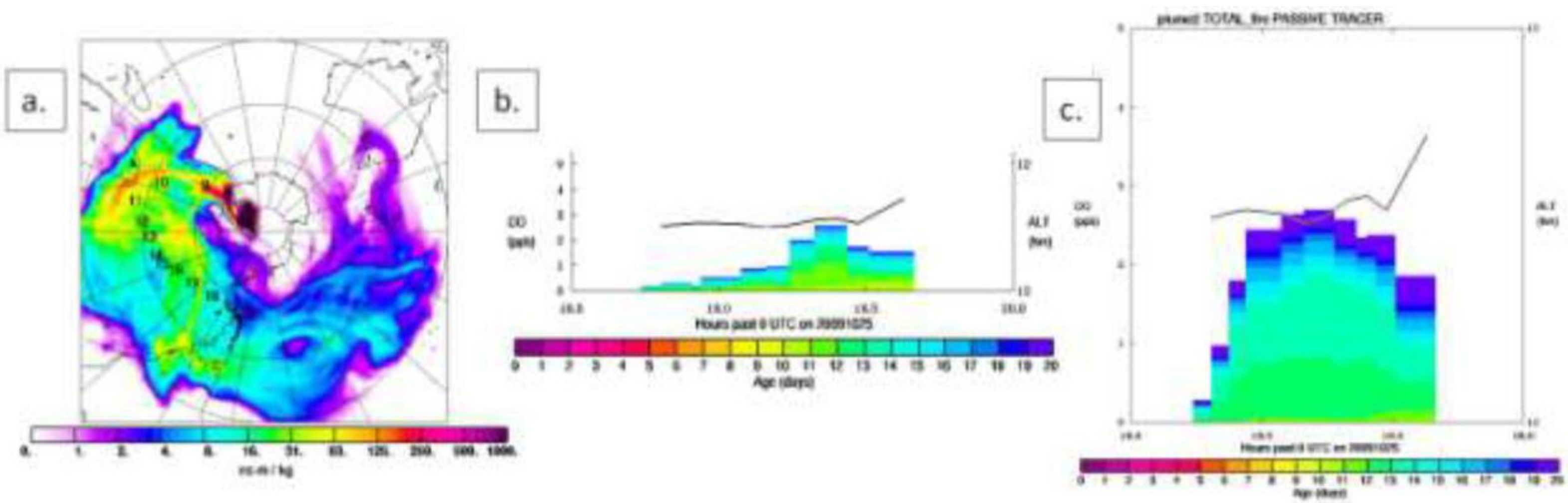


Fig. 4. Simulated daily distribution of IEPOX in the planetary boundary layer during the Northern Hemisphere summer (**A**) and winter (**B**). IEPOX seasonal cycle mirrors the isoprene emissions. The mixing ratio of IEPOX is higher in the tropics than in other isoprene production regions in the northern mid-latitudes (e.g., the southeast United States). This reflects the reduction in the yield of IEPOX from isoprene due to anthropogenic emissions of NO.

10. Choose the right format and resolution

- Good formats: ps, eps, tiff (for photos)
- Bad formats: gif, jpg
- Plot to the correct published size. Do not enlarge with software.
- Submit high resolution figures for review



11 …

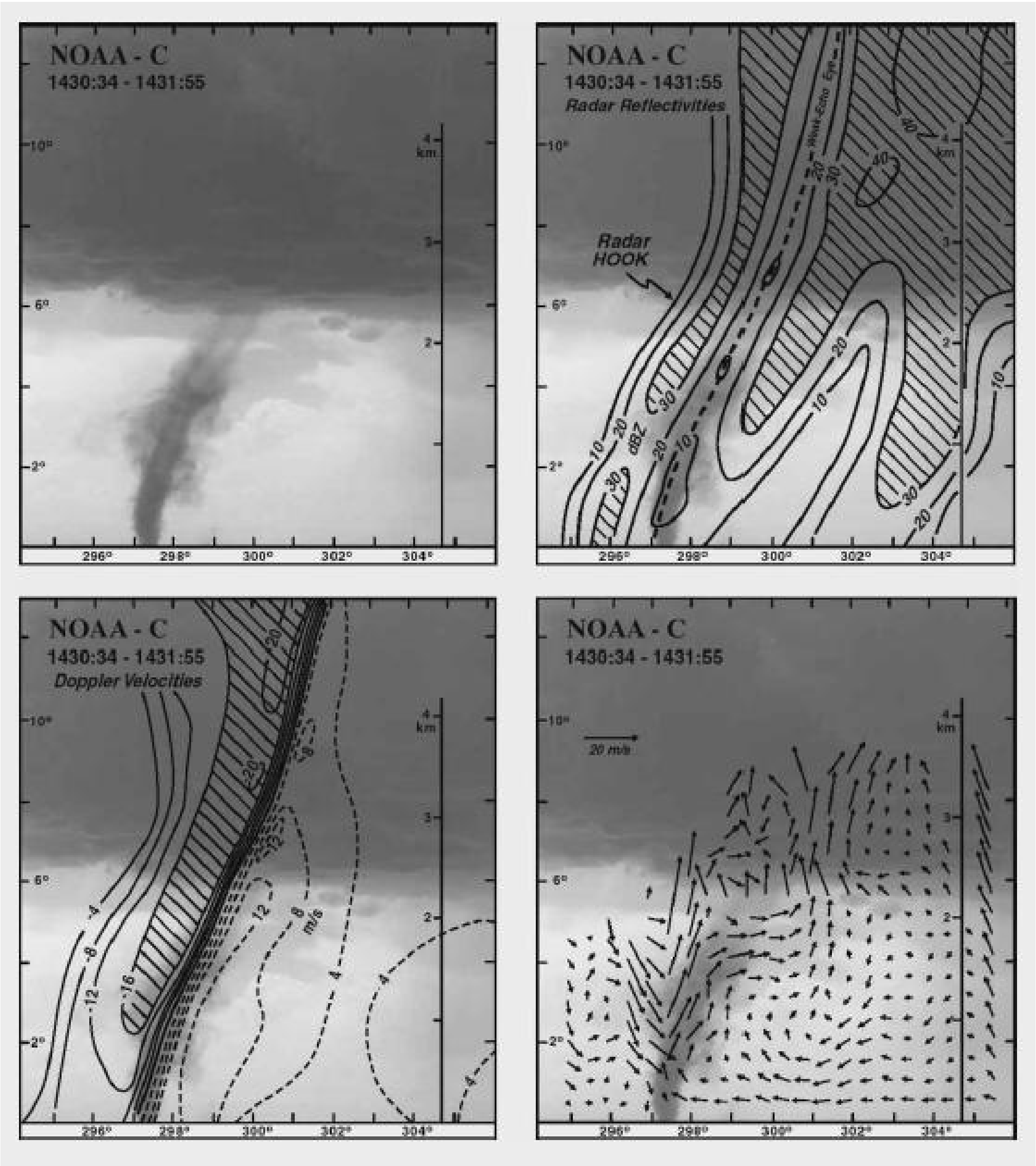
Revise and edit!

没有最好，只有更好

Develop a consistent style

1. Easier to apply your code to new datasets to make similar plots
2. Reduce mistakes and bad plots
3. Easier to put together a consistent presentation in the future
4. Consistently show a good, clear style can be your trademark and your advertisement

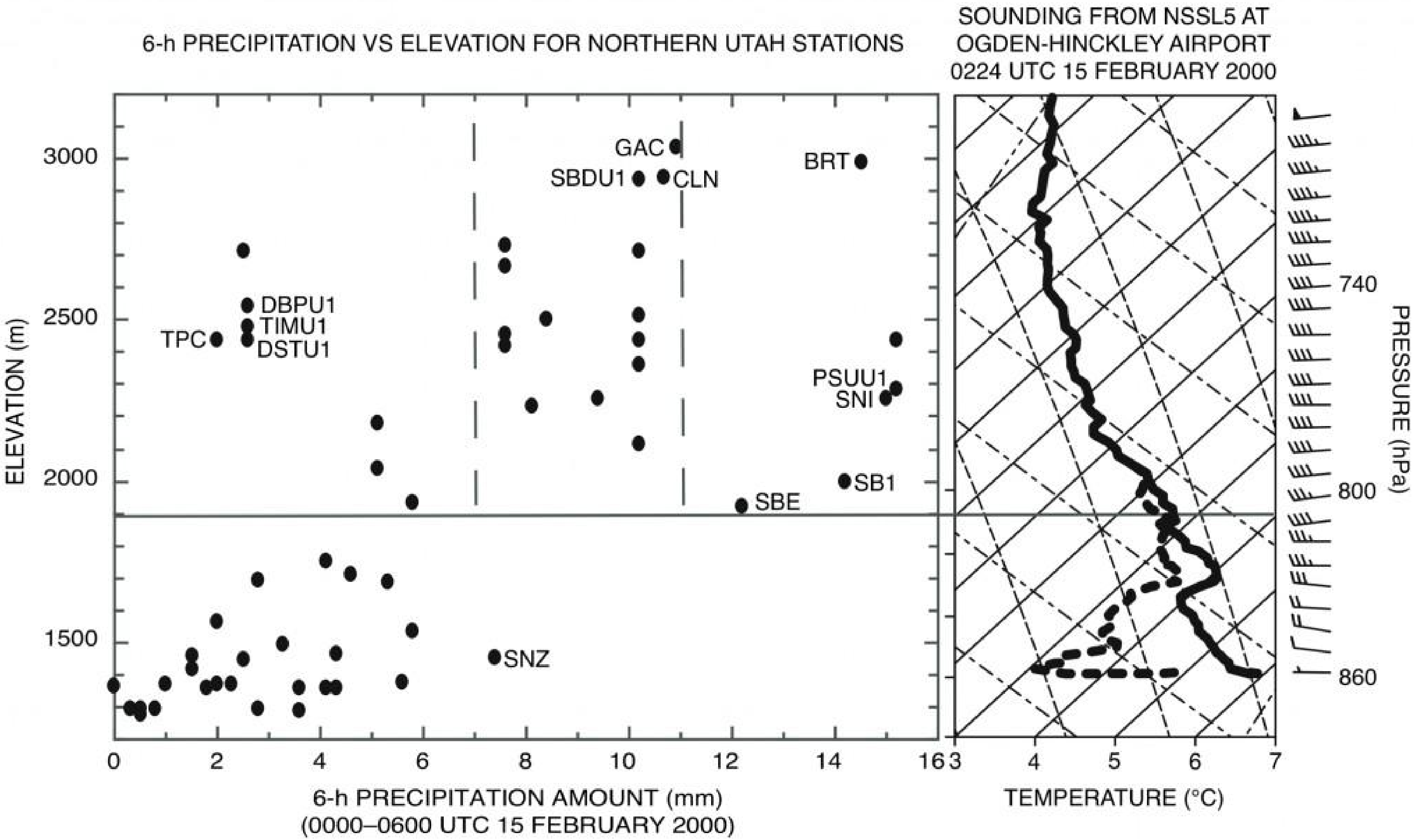
Creative composition (but only if it works!)



Wakimoto and Martner (1992)

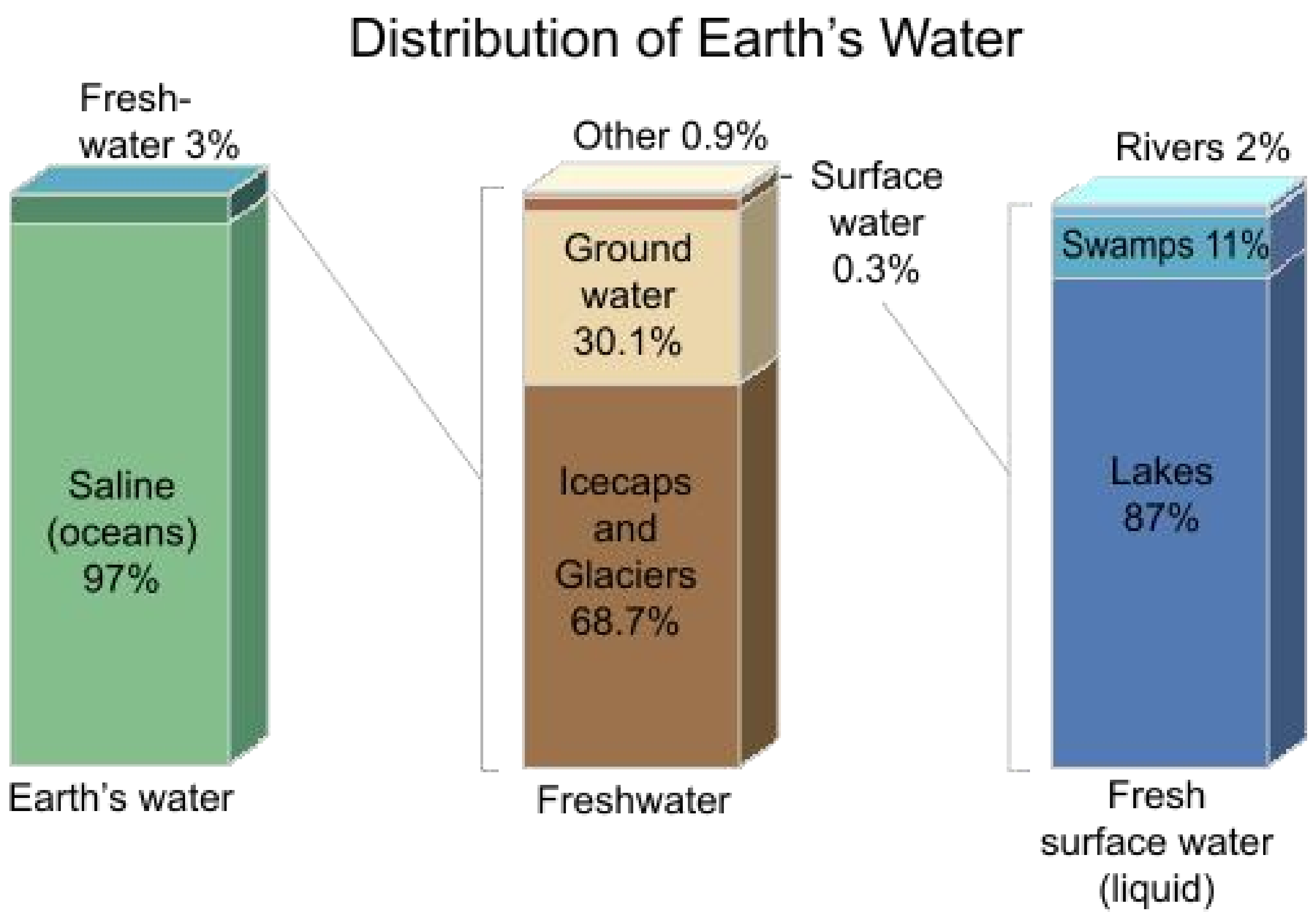
Schultz (2009),
Eloquent Science

Creative composition (but only if it works!)

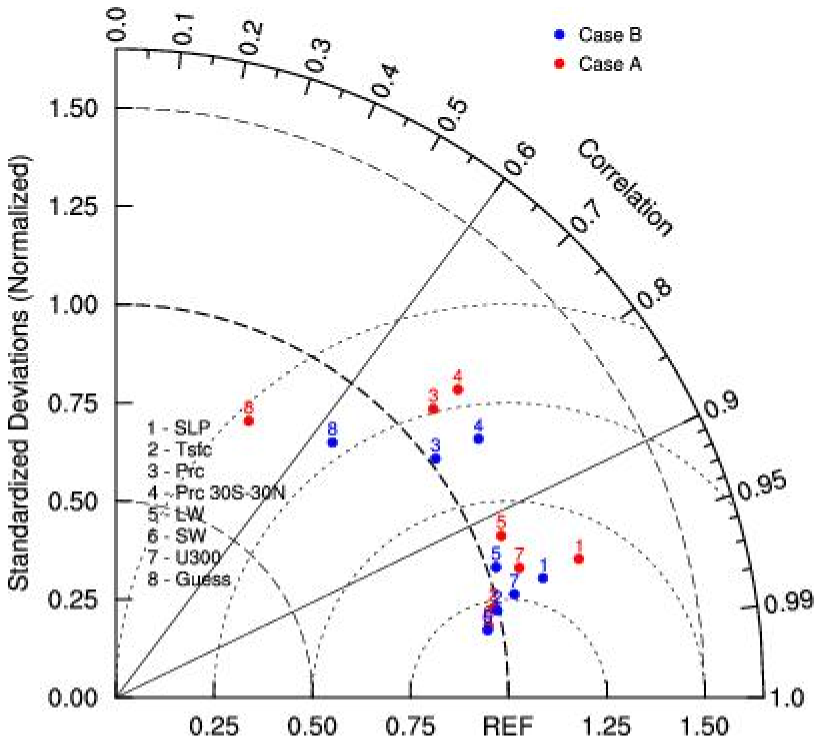


Schultz et al. (2003), *MWR*

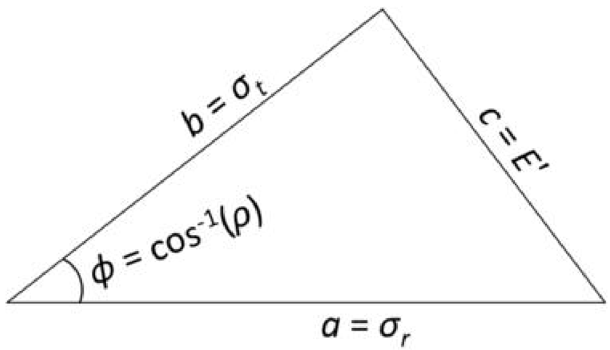
Creative composition (but only if it works!)



Creative composition (but only if it works!)



$$E'^2 = \sigma_r^2 + \sigma_t^2 - 2\sigma_r\sigma_t\rho_r$$

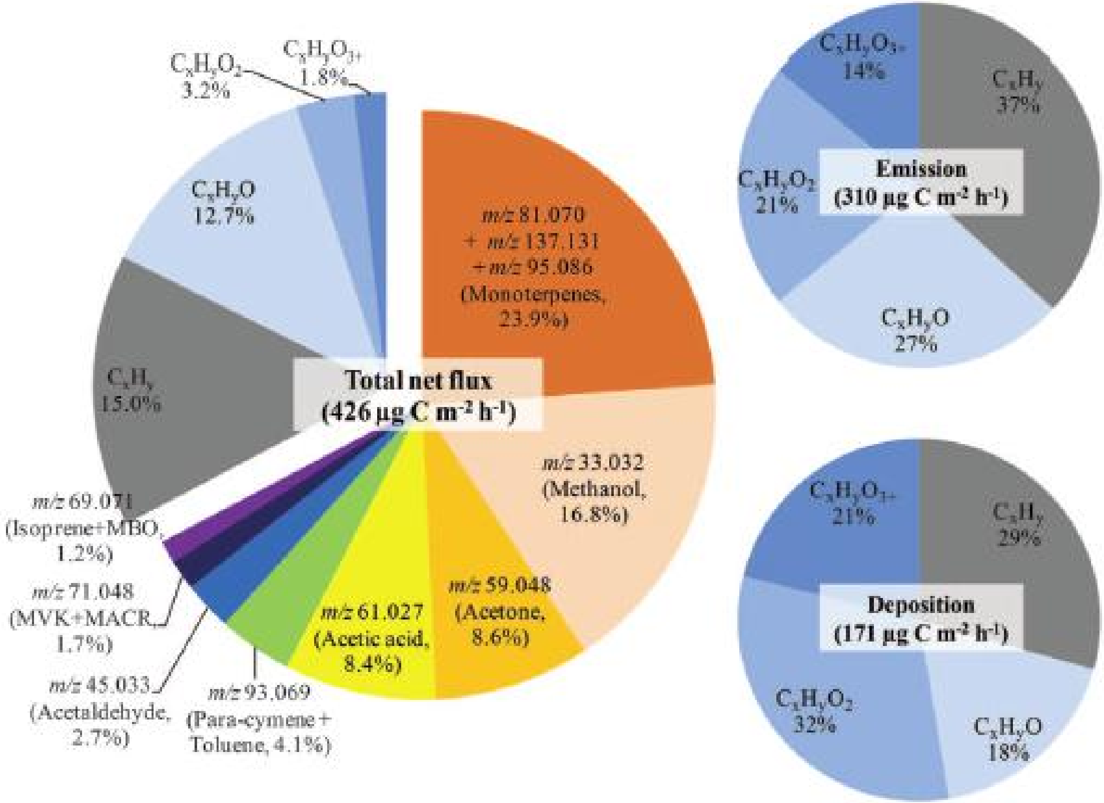


Taylor diagram
http://www.ncl.ucar.edu/Applications/taylor.shtml

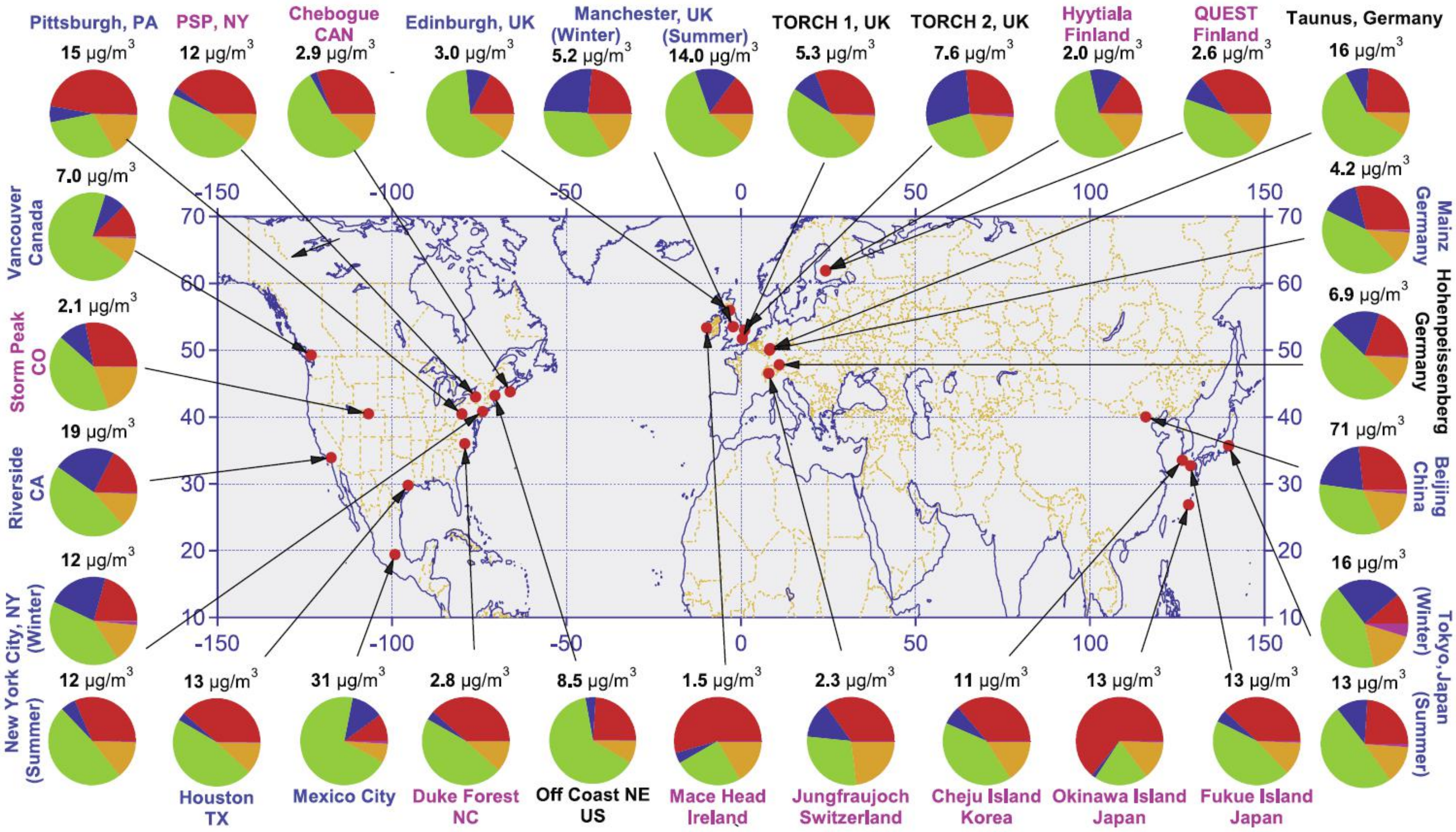
[Taylor, 2001]

No pies!

Fig. 4. Flux contribution by chemical composition. Individual VOC and VOC-group contribution to the total flux for ions to which an empirical formula has been attributed (162 ions) are shown in pie chart on the left. The 10 major masses were specifically identified, and the remaining 152 *m/z* ratios were categorized by number of oxygens in the molecule as C_xH_y, C_xH_yO, $C_xH_yO_2$, and $C_xH_yO_{3+}$. The two pie charts on the right show the contribution of categorized ions to the estimated emission (top) and deposition (bottom). MBO, 2-methyl-3-butene-2-ol.



Creative composition (but only if it works!)



Zhang et al. (2007), *GRL*

Make the key message "scream"!

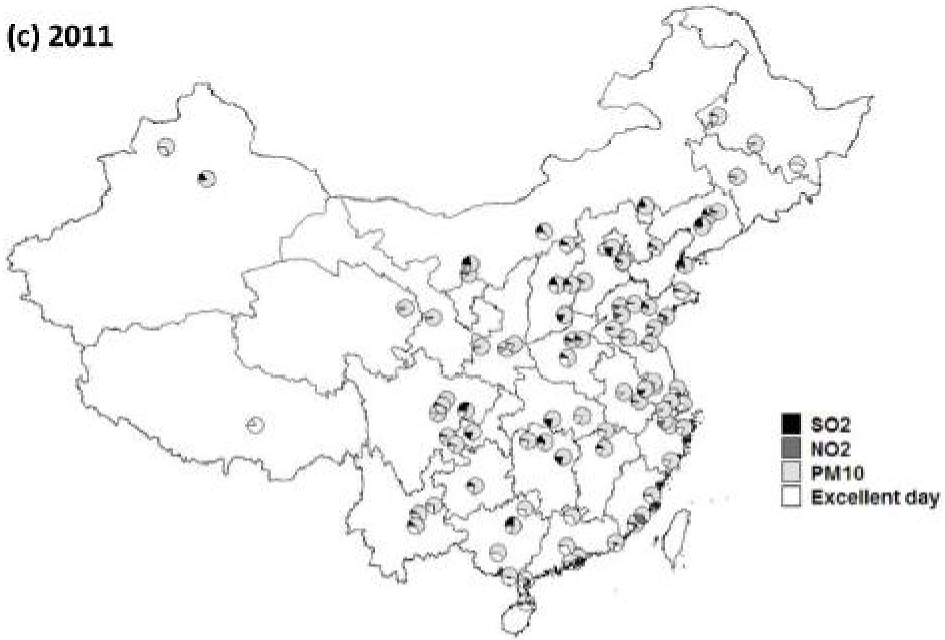
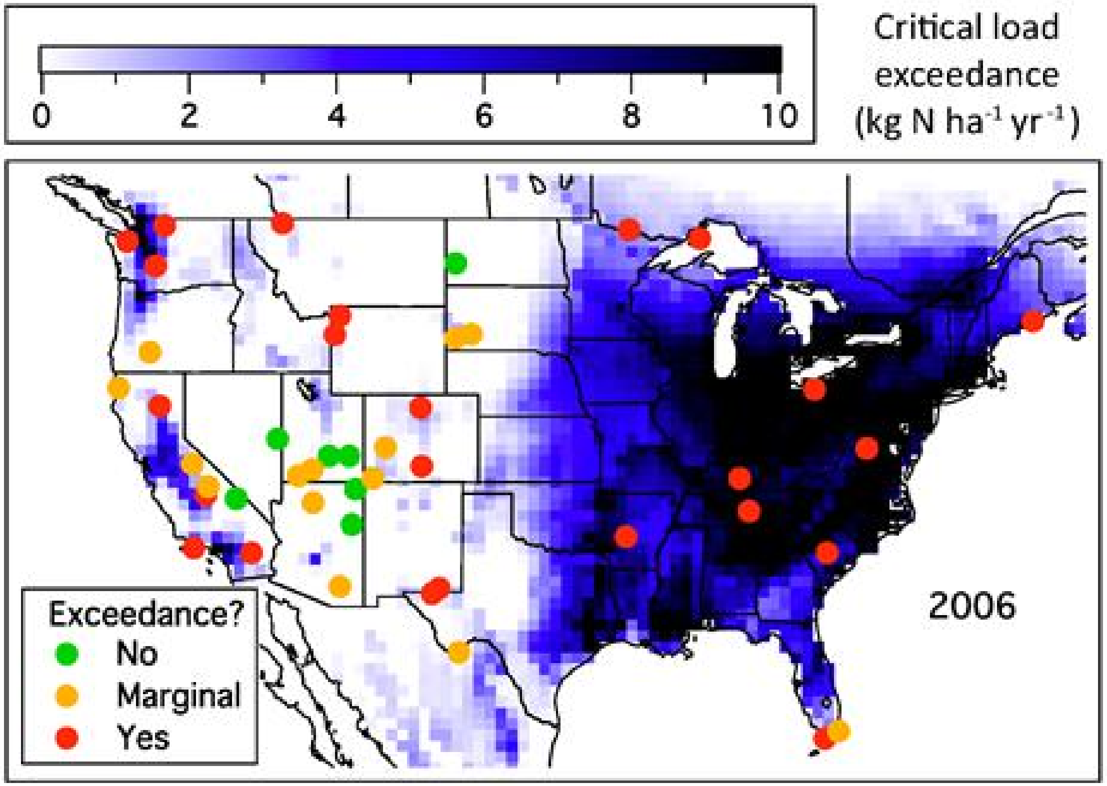


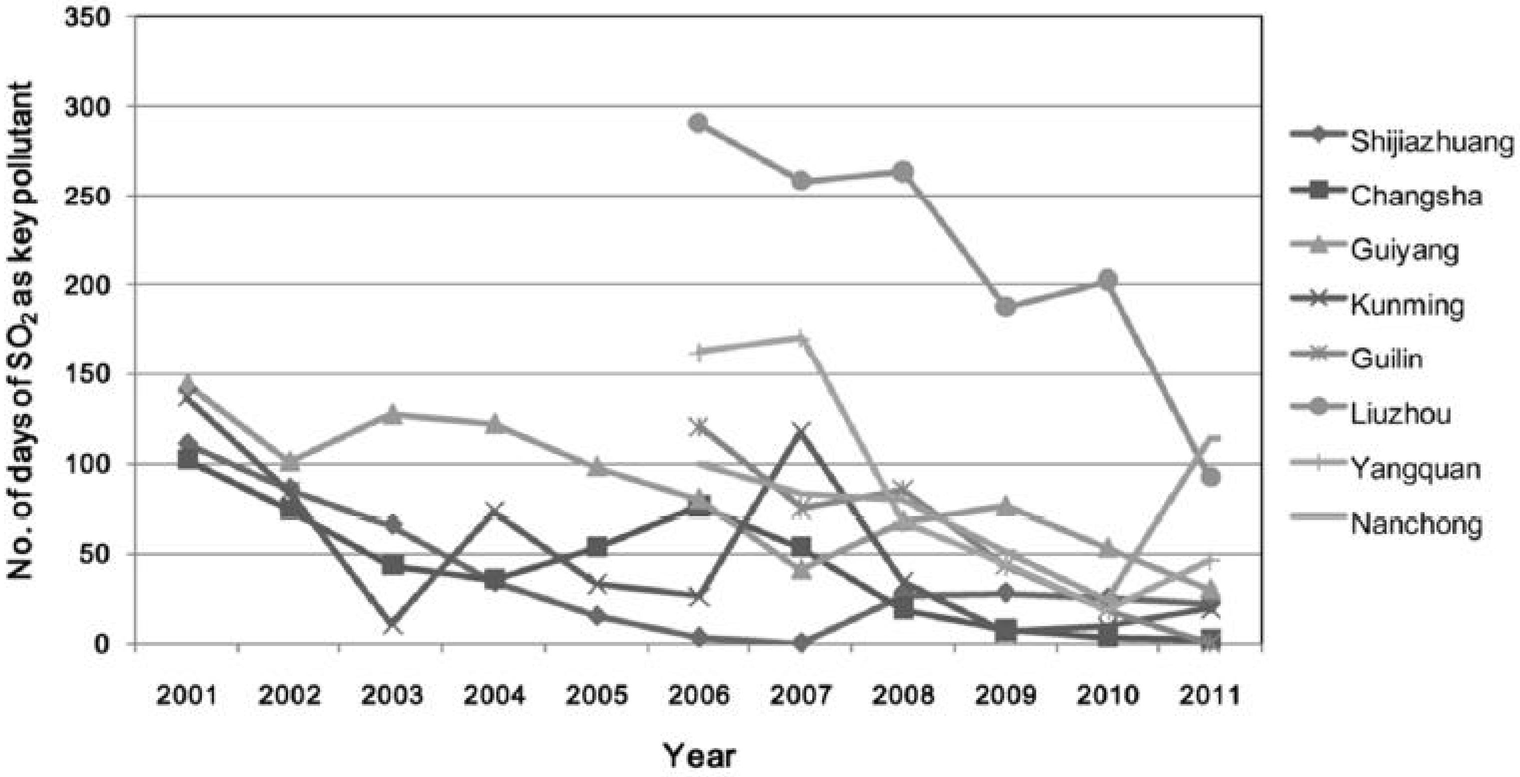
Figure 2. The frequencies of three key pollutants and excellent days in 2001, 2006 and 2011.
234x459mm (300 x 300 DPI)

Creative color use (but only if it works!)

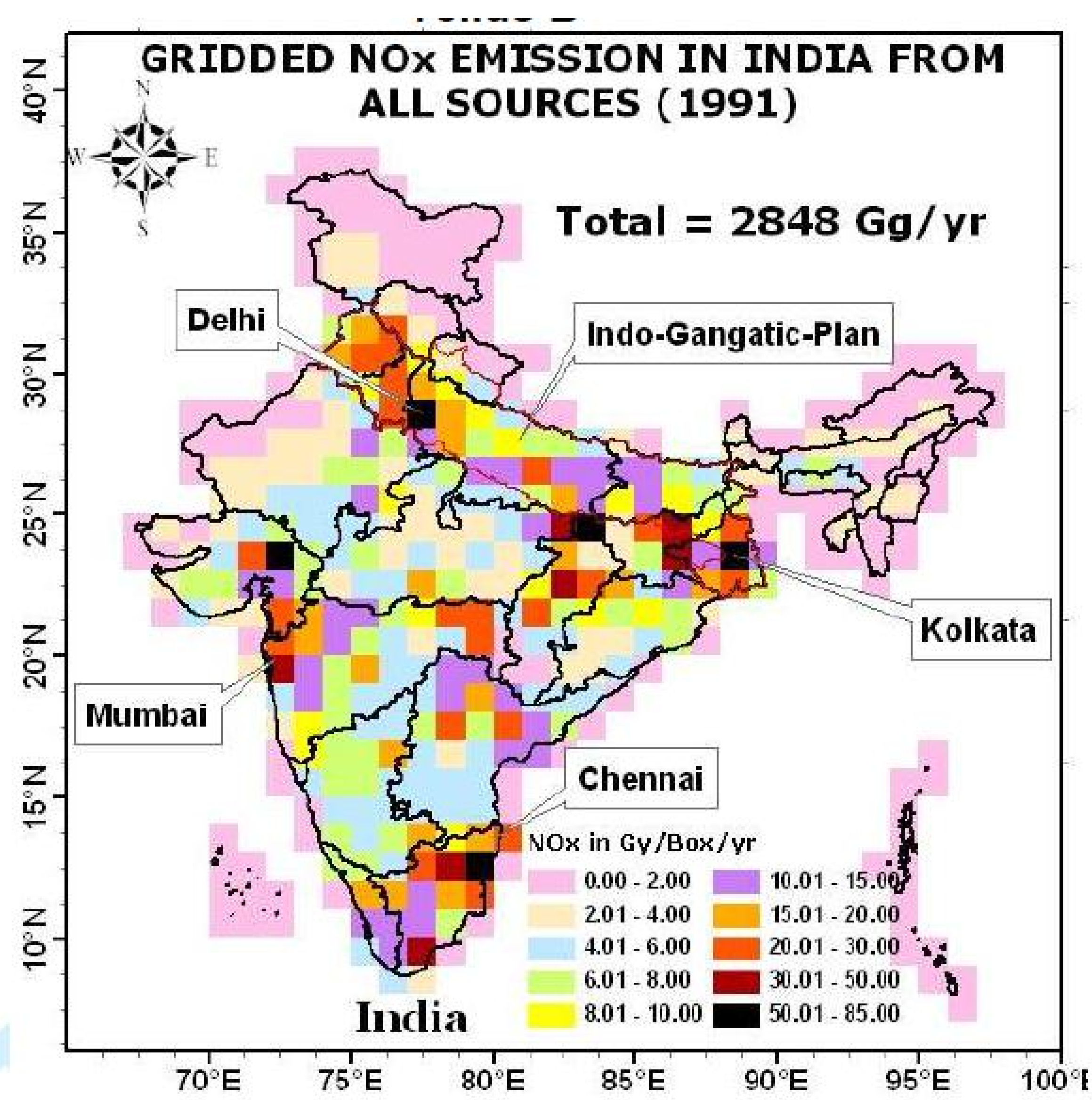


Ellis et al. (2013), *ACP*

Poor examples

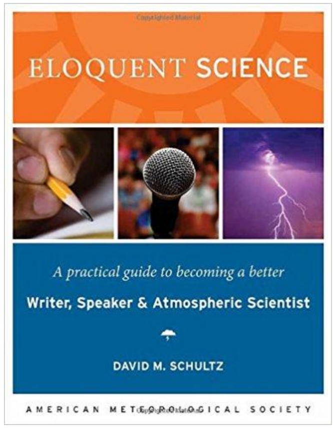


Avoid chartjunk



(Figure: 1)

Resources



- David Schultz's book *Eloquent science* and his blog eloquentscience.com/
- Buckingham (2008), Effective scientific illustrations, ***Lab Times***
- Rolandi et al. (2011), A Brief Guide to Designing Effective Figures for the Scientific Paper, ***Advanced Materials***
- Rougier et al. (2014), Ten simple rules for better figures, ***PLOS Computational Biology***
- PSU College of Earth & Mineral Sciences, Style for students

Work in progress

Thank you!