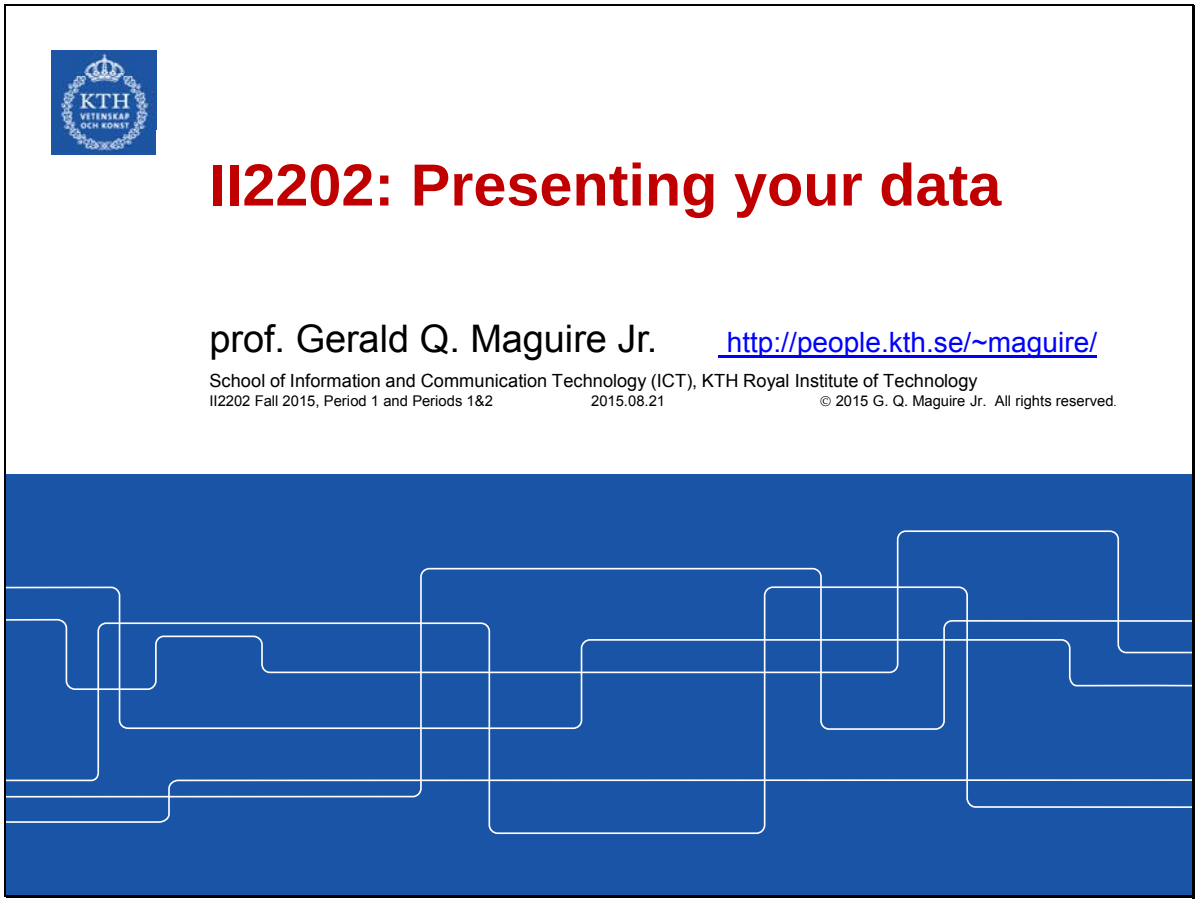
Slide 1



Slide 2

Presenting your data

1. Data visualization for **yourself** – to understand your data, to “listen to your data”, to exploit your data, …
2. **Presenting it to your audience** – so that you can explain what you have found in the data, so that you can facilitate their understanding the data, so that they can make use of your data to build upon it, …

Slide 3

Presenting information with images

"A picture is worth a thousand words."

-- Popular saying

Pictures, graphs, flow charts, UML, state machines, … can convey an enormous amount of information if used well.

Consider "a wink" at a party

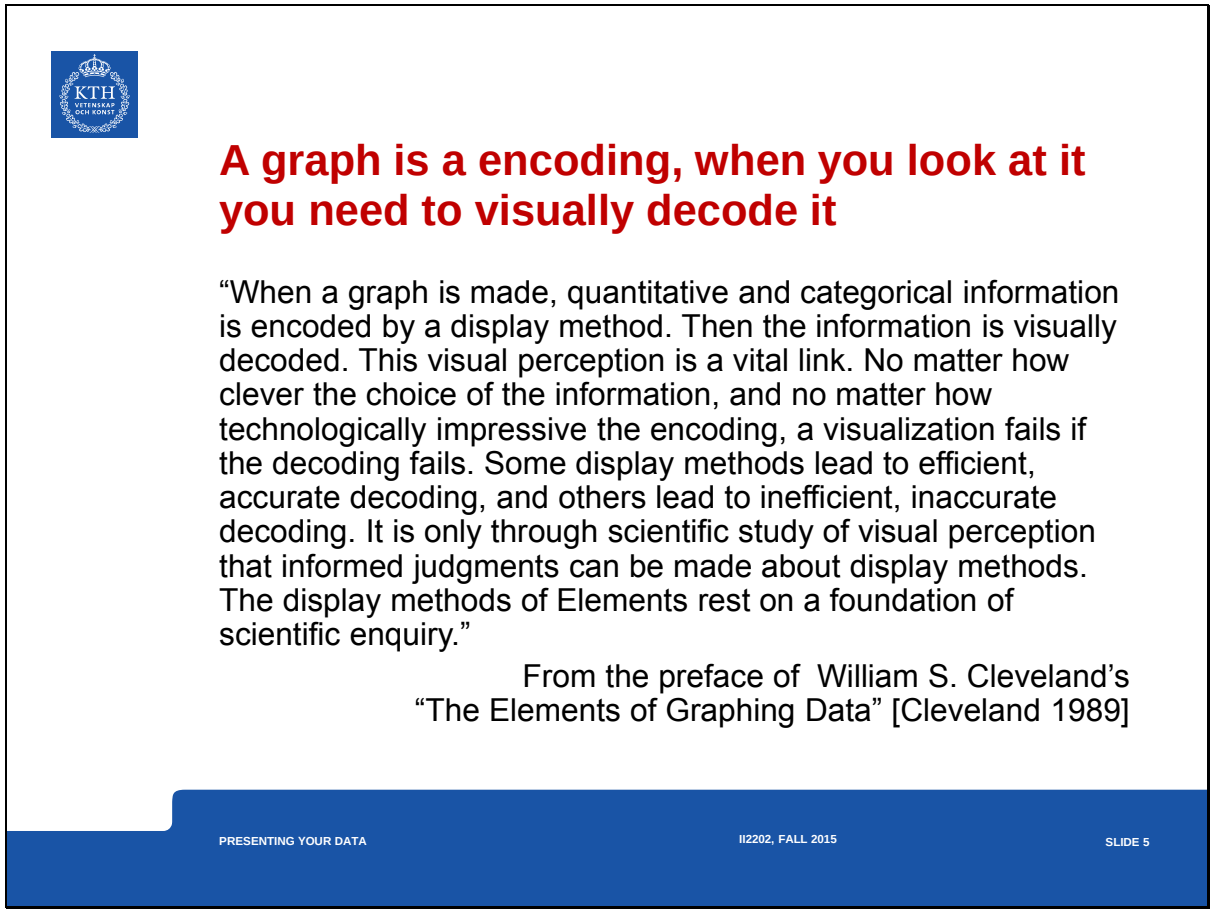
Slide 4

Why use graphical presentations?

- Very compact – you can present at lot of data in a small space – in contrast to a table
- To bring out difference and make comparisons
- To help abstract a general (abstract) “picture” (conception) from the data
- Many people are good at seeing **patterns** in visual scenes
- To achieve clarity and objectivity
- To support your text (i.e., to help you tell your story)

PRESENTING YOUR DATA II2202, FALL 2015 SLIDE 4

Slide 5

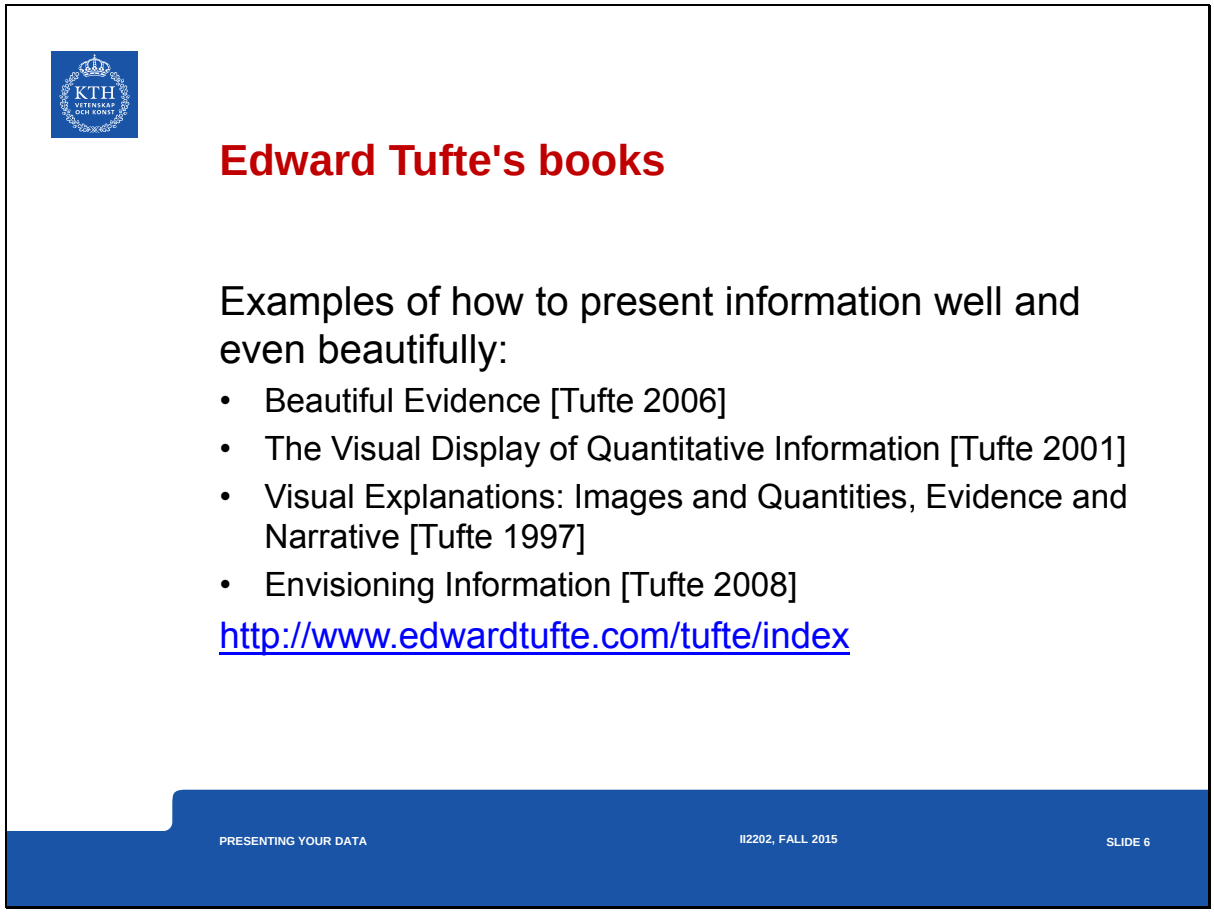


[Cleveland 1989] William S. Cleveland, *The elements of graphing data*, 10.[print.] ed. Monterey, Cal: Wadsworth, 1989, ISBN: 978-0-534-03729-1.

[Cleveland 1993] William S. Cleveland, *Visualizing data*. Murray Hill, N.J. : [Summit, N.J: At&T Bell Laboratories ; Published by Hobart Press, 1993, ISBN: 978-0-9634884-0-4.

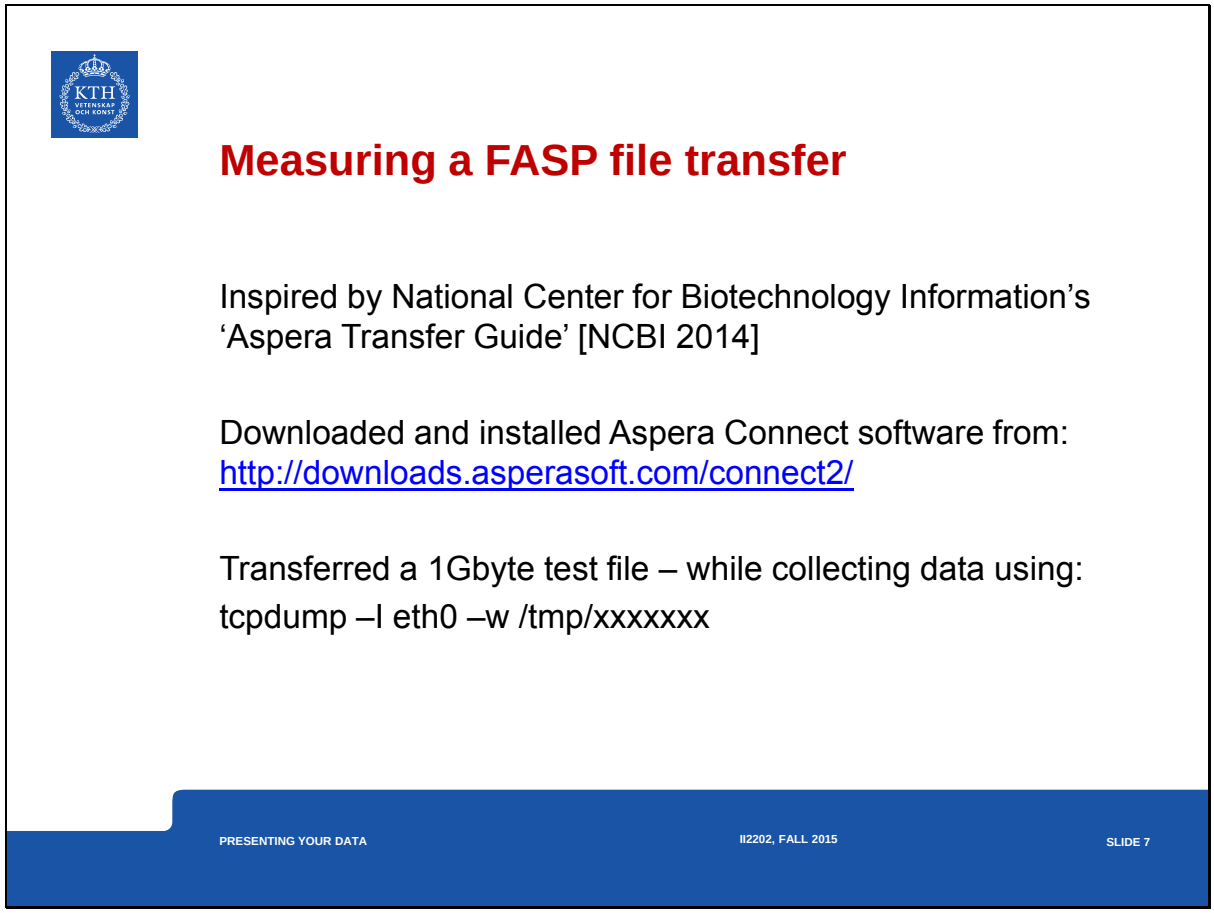
The R code for the figure is the book and the data tables can be found from http://www.stat.purdue.edu/~wsc/visualizing.html

Slide 6



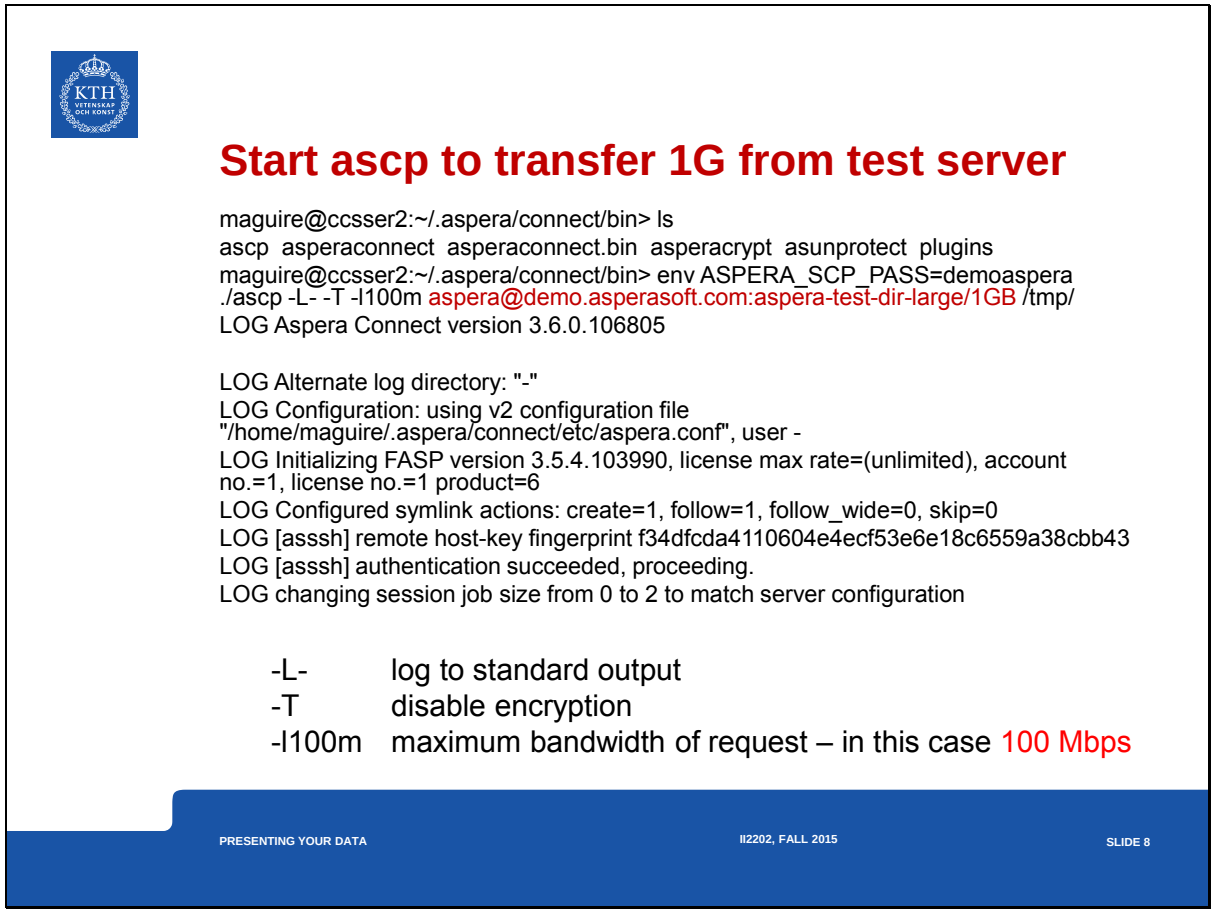
[Tufte 2008] Edward Rolf Tufte, *Envisioning information*, 12. printing. Cheshire, Conn: Graphics Press, 2008, ISBN: 978-0-9613921-1-6.
[Tufte 1997] Edward R. Tufte, *Visual explanations: images and quantities, evidence and narrative*. Cheshire, Conn: Graphics Press, 1997, ISBN: 978-0-9613921-2-3.
[Tufte 2006] Edward R. Tufte, *Beautiful evidence*. Cheshire, Conn: Graphics Press, 2006, ISBN: 978-0-9613921-7-8.
[Tufte 2001] Edward R. Tufte, *The visual display of quantitative information*, 2nd ed. Cheshire, Conn: Graphics Press, 2001, ISBN: 978-0-9613921-4-7.

Slide 7



[NCBI 2014] National Center for Biotechnology Information, 'Aspera Transfer Guide', *National Center for Biotechnology Information, U.S. National Library of Medicine*, 16-Apr-2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK242625/. [Accessed: 18-Aug-2015]

Slide 8



Start ascp to transfer 1G from test server

maguire@ccsser2:~/.aspera/connect/bin> ls
ascp asperaconnect asperaconnect.bin asperacrypt asunprotect plugins
maguire@ccsser2:~/.aspera/connect/bin> env ASPERA_SCP_PASS=demoaspera
./ascp -L- -T -l100m aspera@demo.asperasoft.com:aspera-test-dir-large/1GB /tmp/
LOG Aspera Connect version 3.6.0.106805

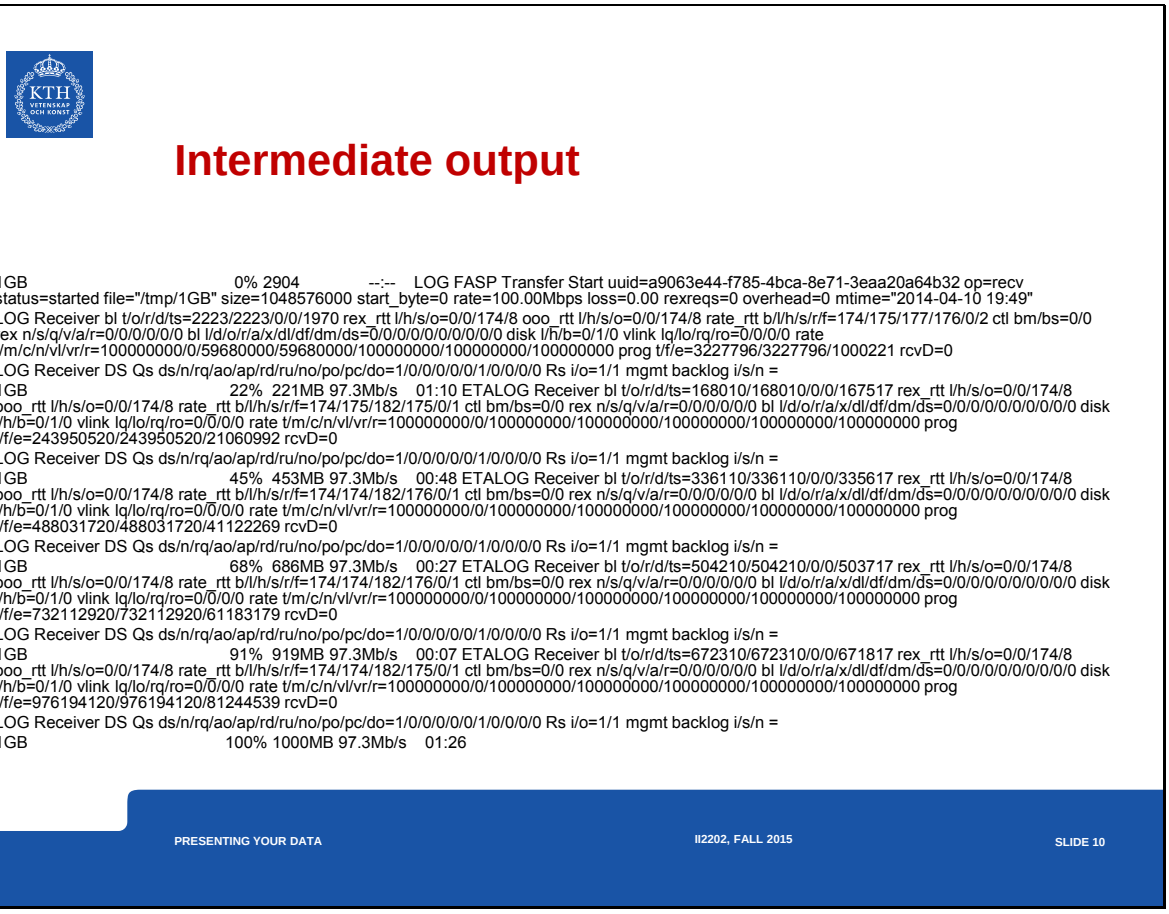
LOG Alternate log directory: "-"
LOG Configuration: using v2 configuration file "/home/maguire/.aspera/connect/etc/aspera.conf", user -
LOG Initializing FASP version 3.5.4.103990, license max rate=(unlimited), account no.=1, license no.=1 product=6
LOG Configured symlink actions: create=1, follow=1, follow_wide=0, skip=0
LOG [asssh] remote host-key fingerprint f34dfcda4110604e4ecf53e6e18c6559a38cbb43
LOG [asssh] authentication succeeded, proceeding.
LOG changing session job size from 0 to 2 to match server configuration

-L-	log to standard output
-T	disable encryption
-l100m	maximum bandwidth of request – in this case 100 Mbps

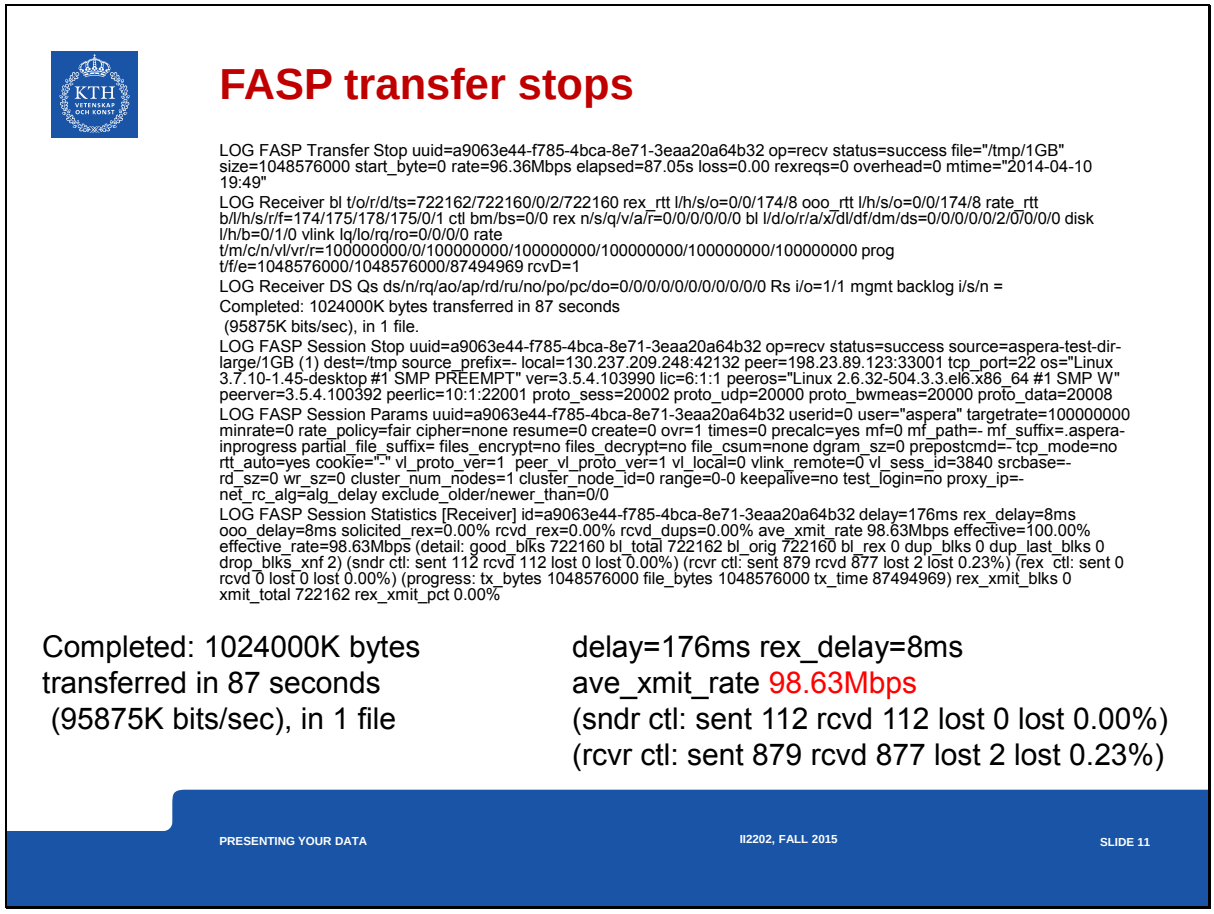
PRESENTING YOUR DATA II2202, FALL 2015 SLIDE 8

Slide 9

Slide 10



Slide 11



Slide 12

Final transfer statistics

```
LOG ======= File Transfer statistics =======
LOG ------- Source statistics -------
LOG Source argument scans attempted           :         1
LOG - Source argument scans completed         :         1
LOG Source path scans attempted               :         1
LOG - Source path scans failed                :         0
LOG - Source path scans skipped since irregular :       0
LOG - Source path scans excluded              :         0
LOG - Source directory scans completed        :         0
LOG - Source file scans completed             :         1
LOG Source directory creates attempted        :         0
LOG - Source directory creates failed         :         0
LOG - Source directory created or existed     :         0
LOG Source file transfers attempted           :         1
LOG - Source file transfers failed            :         0
LOG - Source file transfers passed            :         1
LOG - Source file transfers skipped           :         0
LOG Source bytes transferred                  : 1048576000
LOG ======= end File Transfer statistics =======
```

PRESENTING YOUR DATA II2202, FALL 2015 SLIDE 12

Slide 13

Wireshark: UDP conversion

client server

A: 130.237.209.248:42132 ↔ B: 198.23.89.123:33001

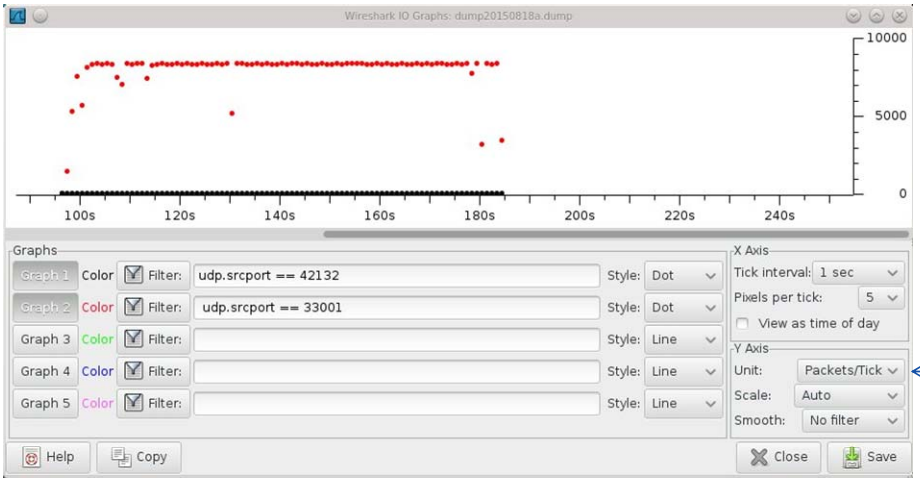
Packets: 703,728 Bytes: 1,058,302,232

	Packets	Bytes	bps
A→B	961	88518	8,058.85
A←B	702,767	1,058,213,714	96,341,791.88

File size = 1,048,576,000 bytes

Slide 14

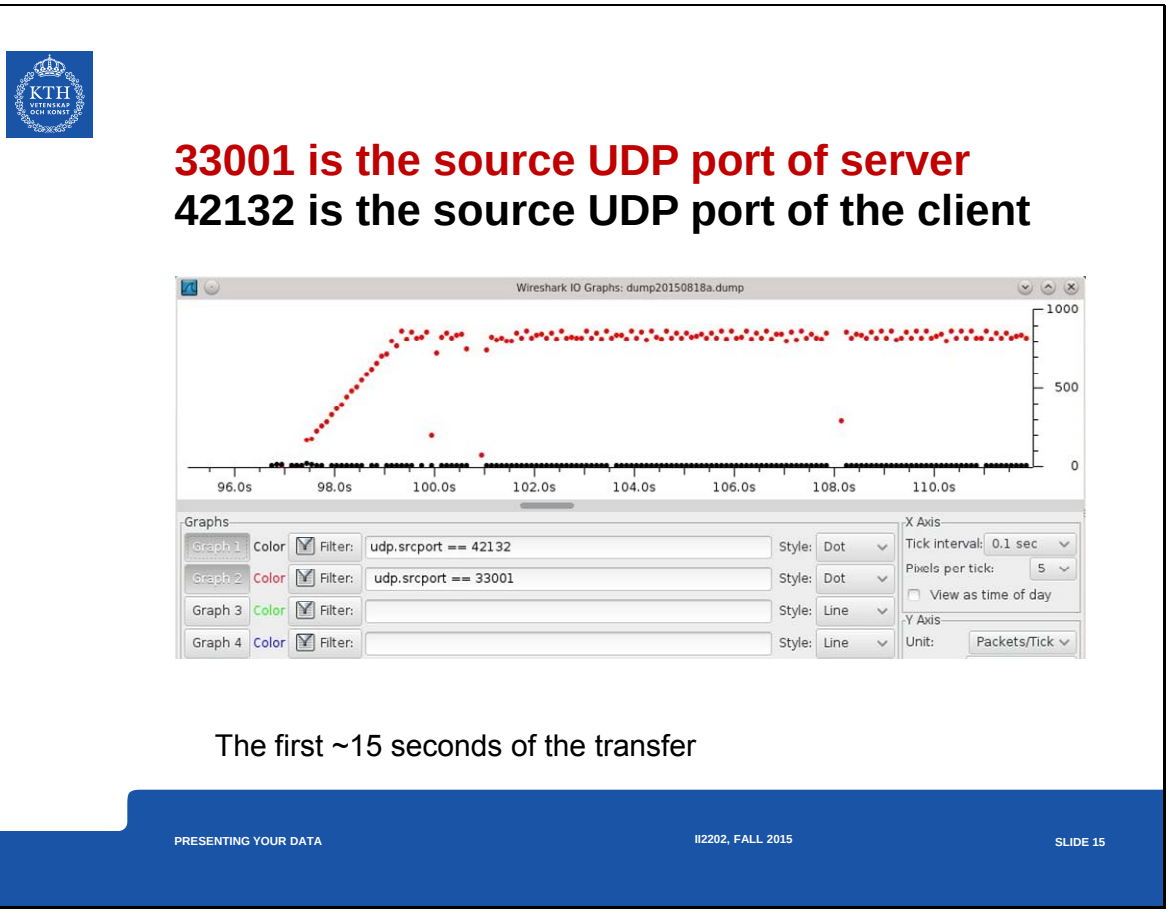
33001 is the source UDP port of server
42132 is the source UDP port of the client



← Packet rate

PRESENTING YOUR DATA II2202, FALL 2015 SLIDE 14

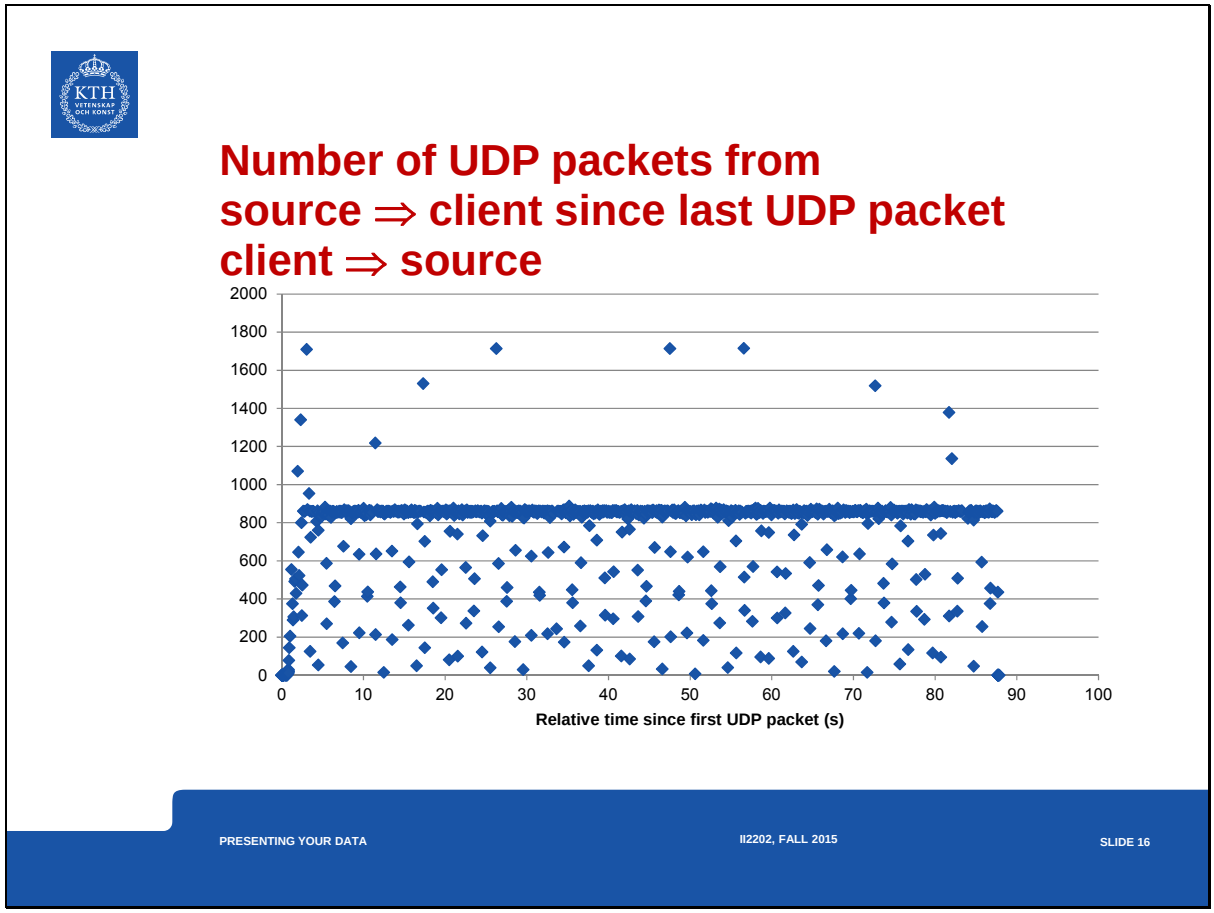
Slide 15



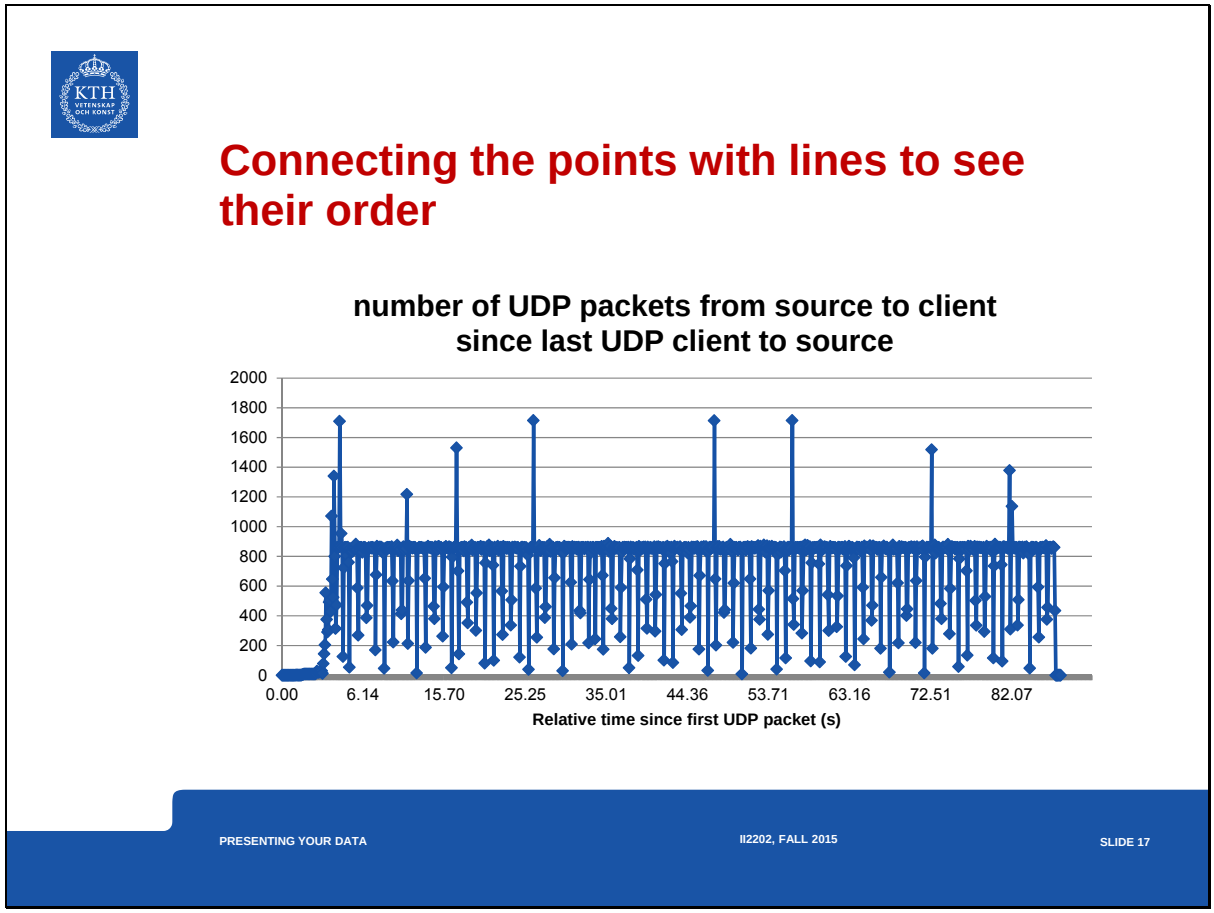
33001 is the source UDP port of server
42132 is the source UDP port of the client

The first ~15 seconds of the transfer

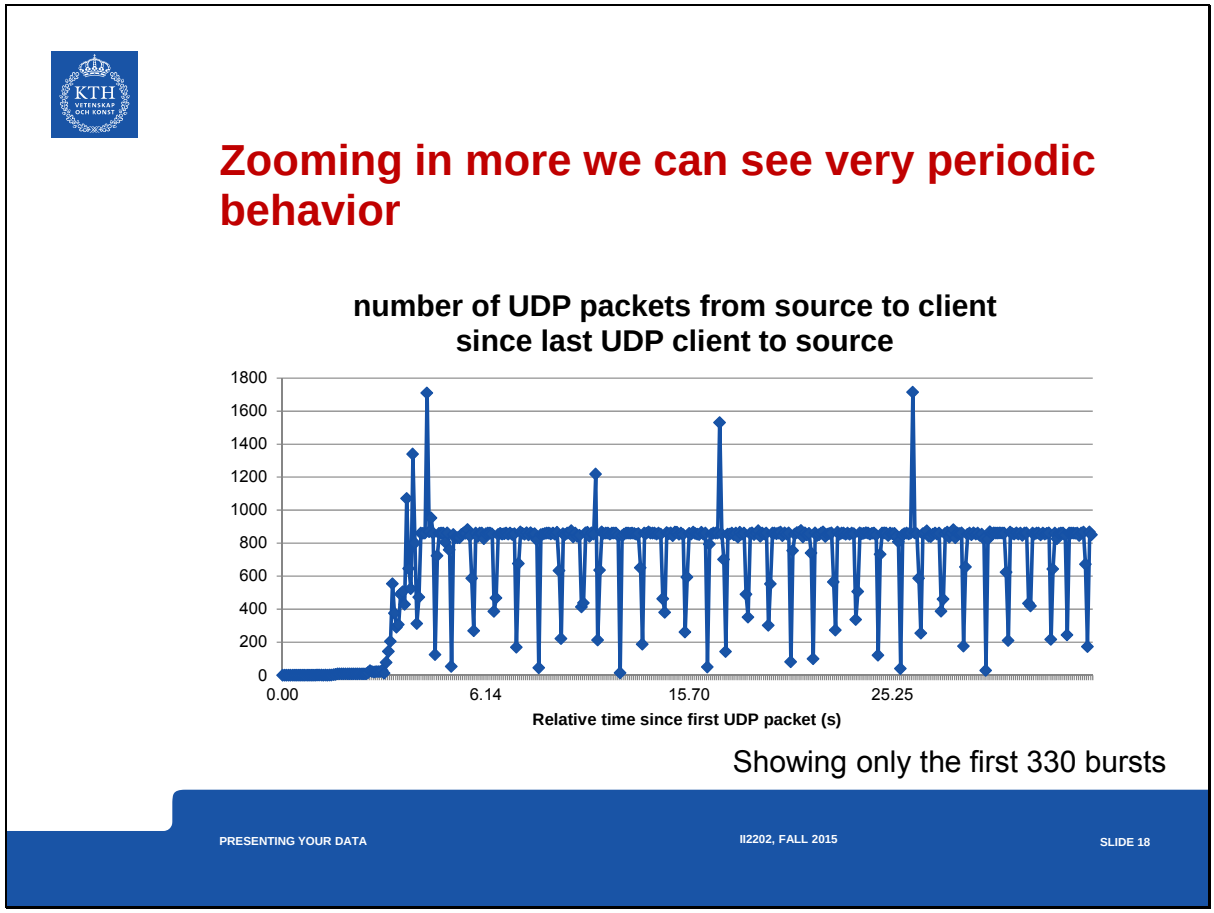
Slide 16



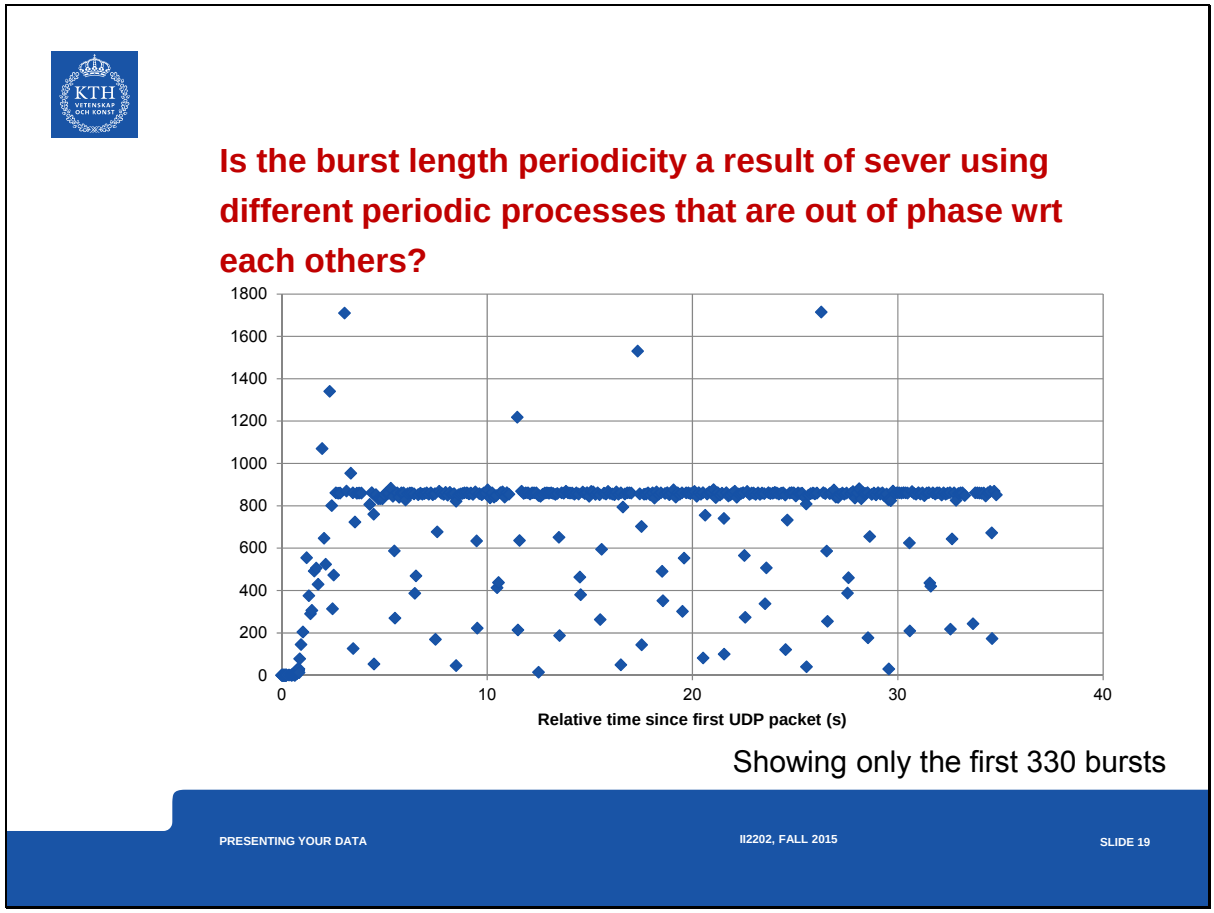
Slide 17



Slide 18



Slide 19



Slide 20

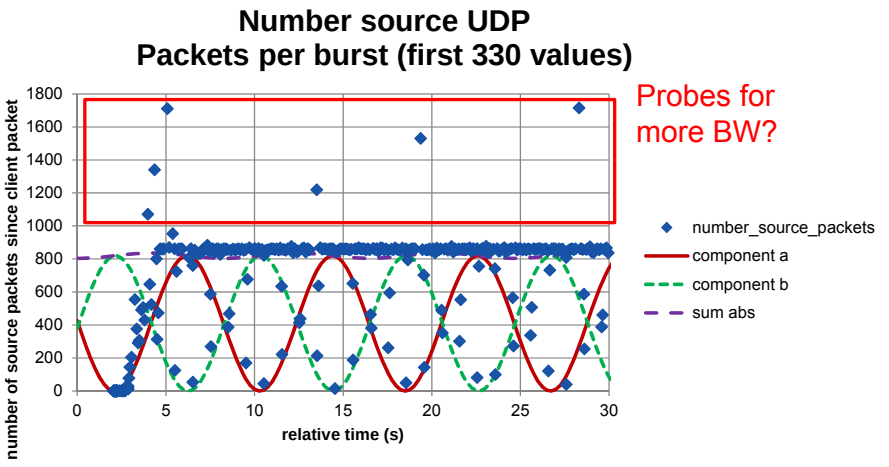
Two sinograms – out of phase with each other – shaping the short burst lengths?

scale	410	410
v_offset	410	410
offset	4.05	8.1
frequency	0.121951	0.121951

1/8.2

median=857

963 bursts in ~90 s

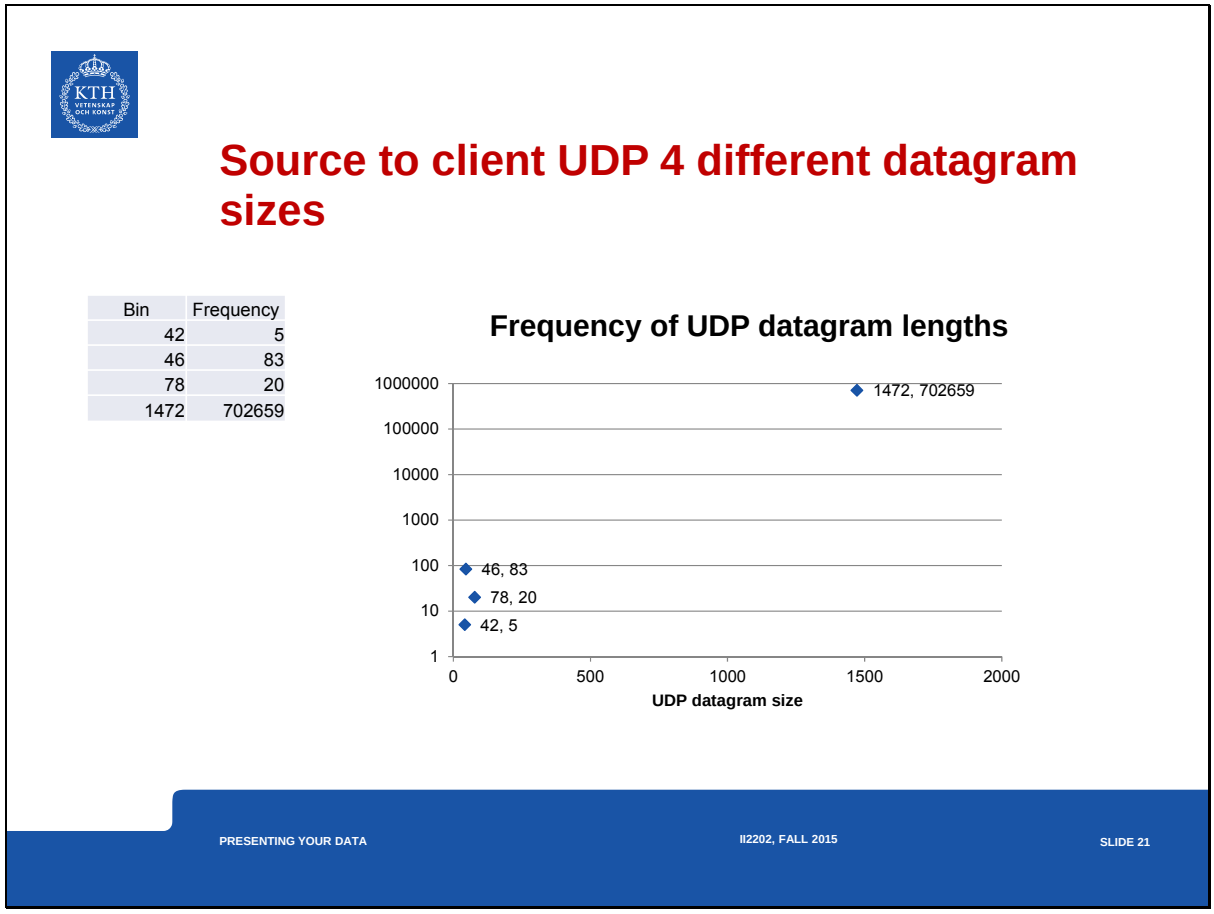


Probes for more BW?

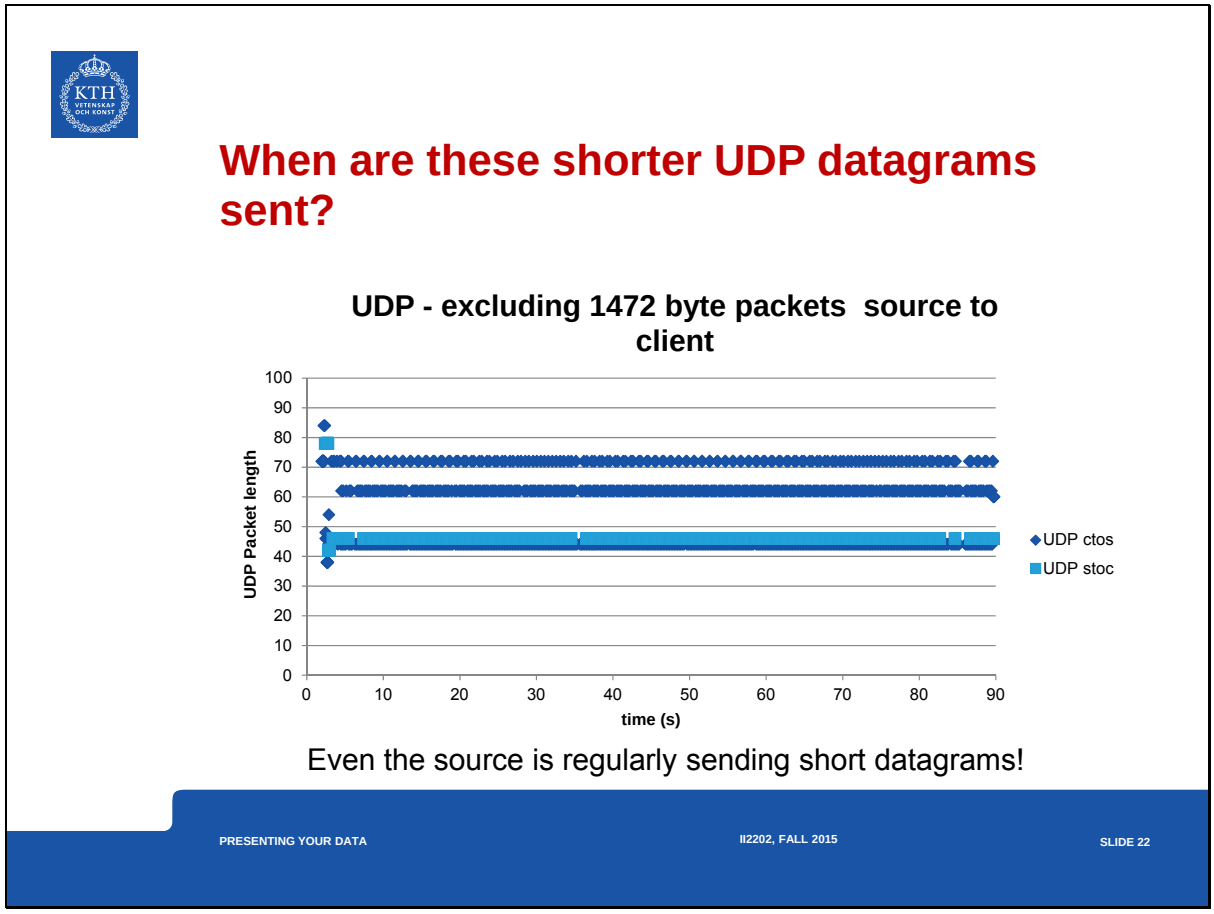
Shorter bursts to detect congestion or decrease in BW?

PRESENTING YOUR DATA II2202, FALL 2015 SLIDE 20

Slide 21



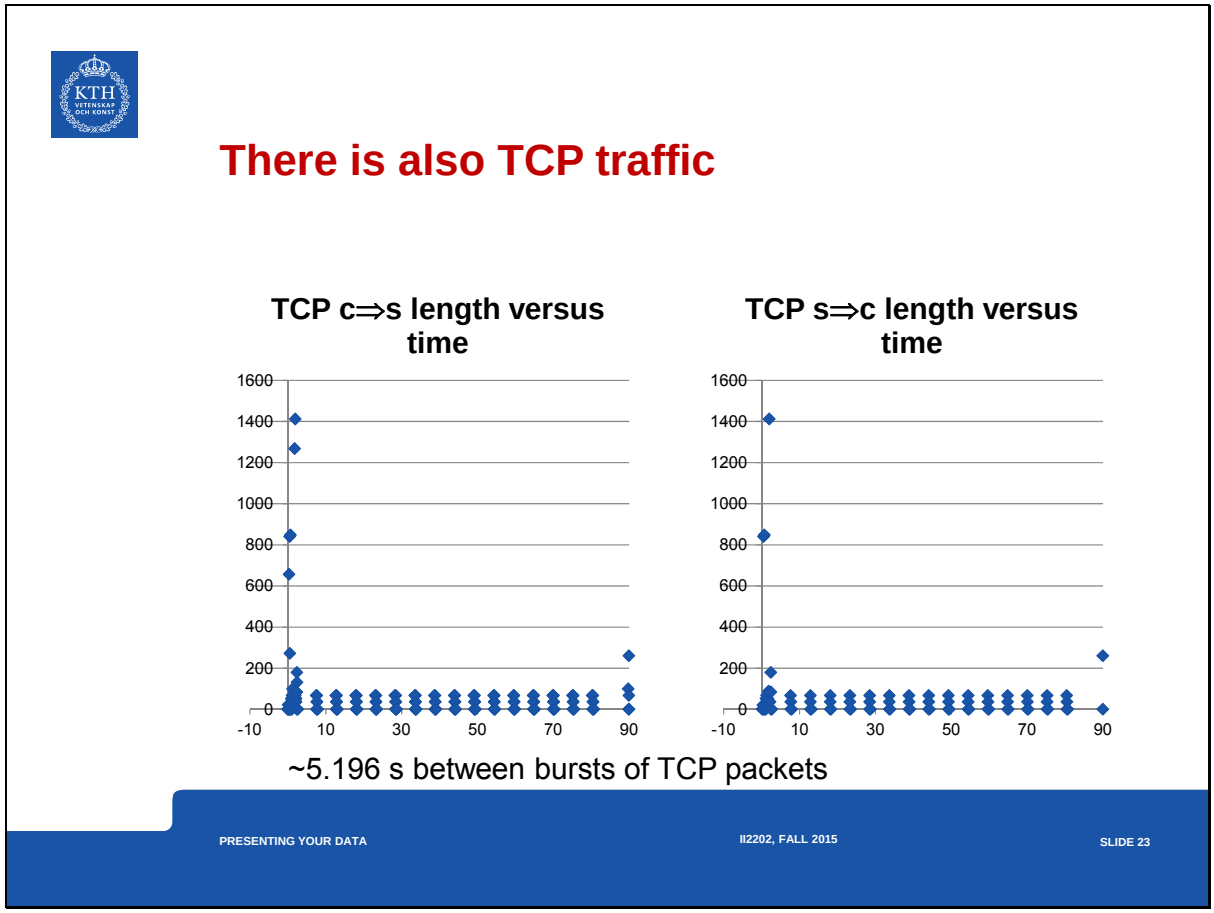
Slide 22



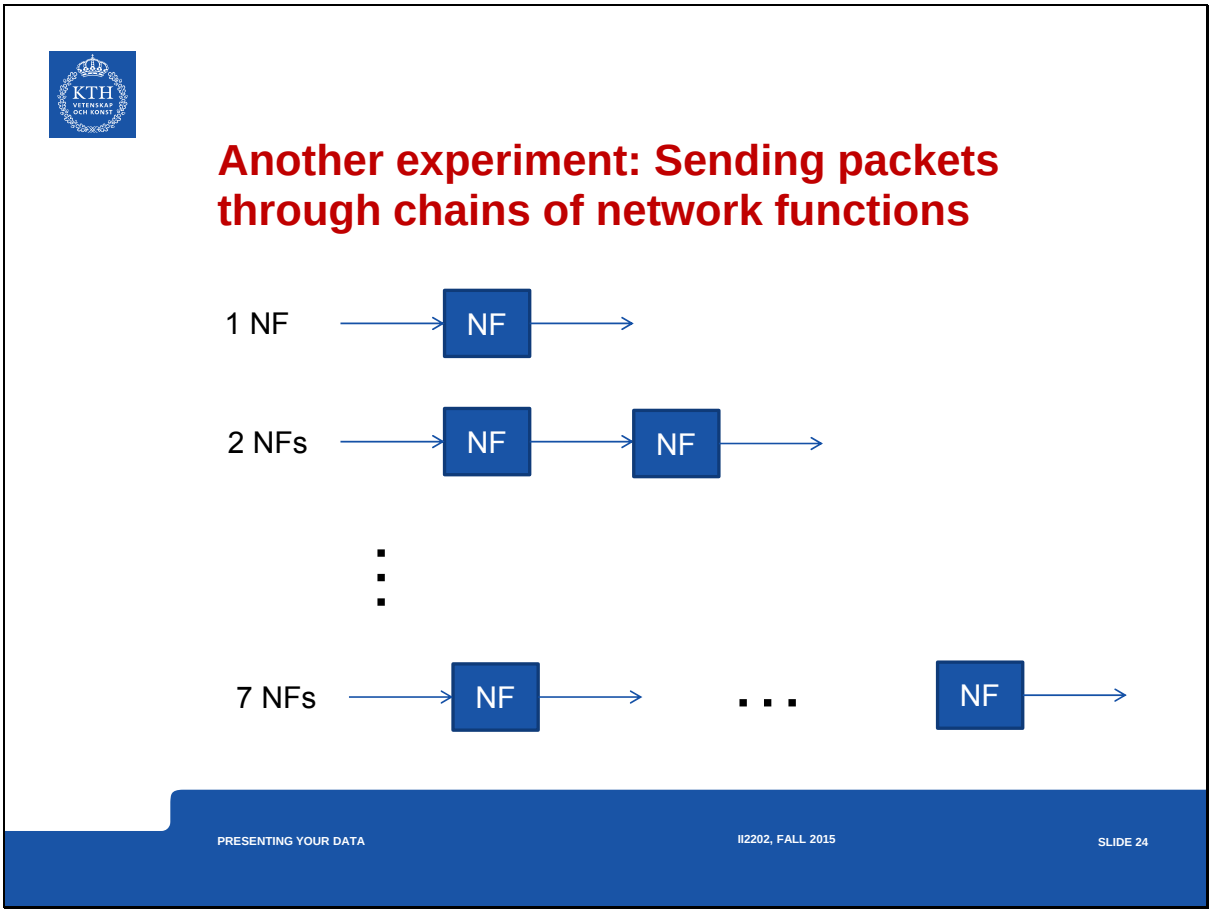
When are these shorter UDP datagrams sent?

Even the source is regularly sending short datagrams!

Slide 23



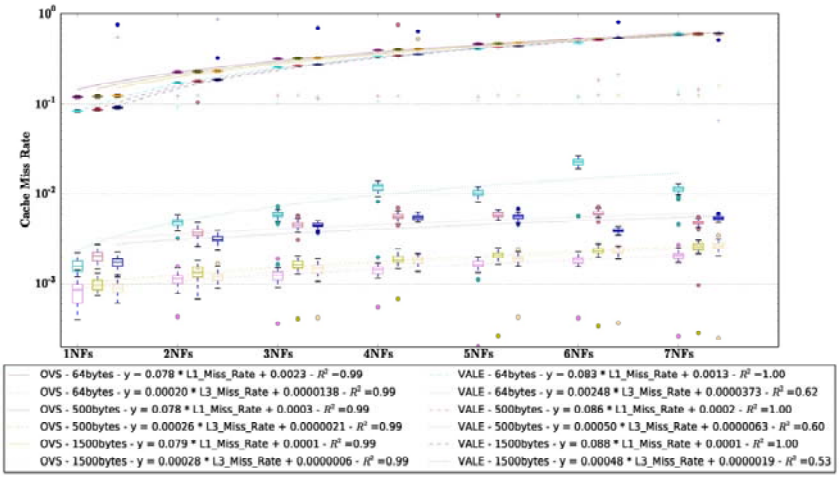
Slide 24



Slide 25

Importance of visualizing outliers

Cache miss rates as a function of the length of a chain of network functions



Data provided by Georgios Katsikas, doctoral student at Network Systems Lab (NSL), CoS

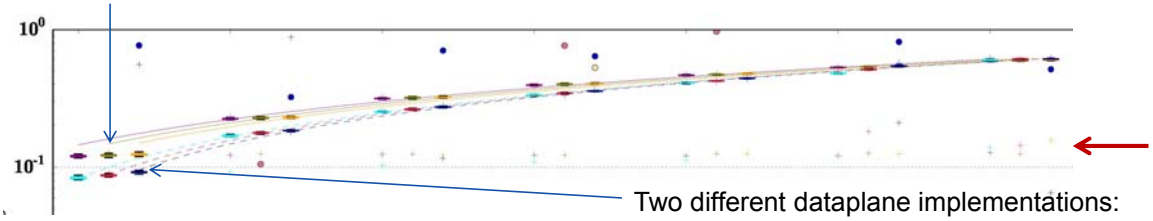
PRESENTING YOUR DATA II2202, FALL 2015 SLIDE 25

Slide 26

Importance of visualizing outliers

Zooming in on L1 cache miss rates as a function of the length of a chain

Three different lengths of packets (64, 500, 1500)



Two different dataplane implementations:
Open vSwitch (OVS) and netmap/VALE

OVS - 64bytes - y = 0.078 * L1_Miss_Rate + 0.0023 - R^2 =0.99	VALE - 64bytes - y = 0.083 * L1_Miss_Rate + 0.0013 - R^2 =1.00
OVS - 64bytes - y = 0.00020 * L3_Miss_Rate + 0.0000138 - R^2 =0.99	VALE - 64bytes - y = 0.00248 * L3_Miss_Rate + 0.0000373 - R^2 =0.62
OVS - 500bytes - y = 0.078 * L1_Miss_Rate + 0.0003 - R^2 =0.99	VALE - 500bytes - y = 0.086 * L1_Miss_Rate + 0.0002 - R^2 =1.00
OVS - 500bytes - y = 0.00026 * L3_Miss_Rate + 0.0000021 - R^2 =0.99	VALE - 500bytes - y = 0.00050 * L3_Miss_Rate + 0.0000063 - R^2 =0.60
OVS - 1500bytes - y = 0.079 * L1_Miss_Rate + 0.0001 - R^2 =0.99	VALE - 1500bytes - y = 0.088 * L1_Miss_Rate + 0.0001 - R^2 =1.00
OVS - 1500bytes - y = 0.00028 * L3_Miss_Rate + 0.0000006 - R^2 =0.99	VALE - 1500bytes - y = 0.00048 * L3_Miss_Rate + 0.0000019 - R^2 =0.53

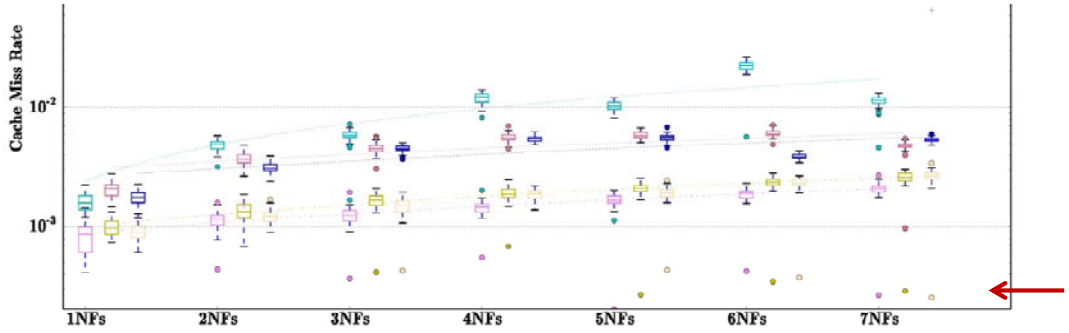
Data provided by Georgios Katsikas, doctoral student at Network Systems Lab (NSL), CoS

PRESENTING YOUR DATA II2202, FALL 2015 SLIDE 26

Slide 27

Importance of outliers

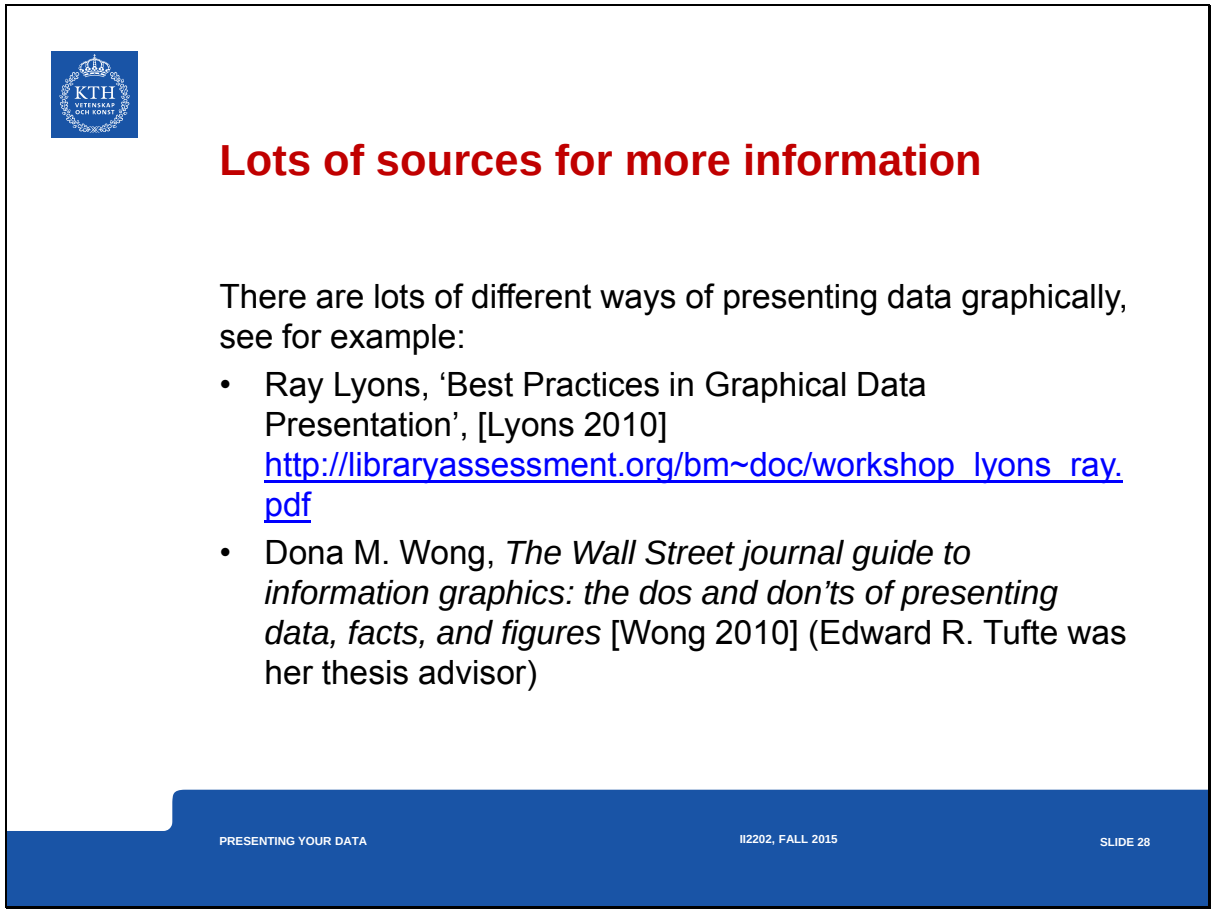
Zooming in on the lowest L3 miss rates – shows there are some that experience an order of magnitude lower cache miss rates. Why?



Data provided by Georgios Katsikas, doctoral student at Network Systems Lab (NSL), CoS

PRESENTING YOUR DATA II2202, FALL 2015 SLIDE 27

Slide 28



[Lyons 2010] Ray Lyons, 'Best Practices in Graphical Data Presentation', Baltimore, MD, USA, 25-Oct-2010 [Online]. Available: http://libraryassessment.org/bm~doc/workshop_lyons_ray.pdf. [Accessed: 18-Aug-2015]

[Wong 2010] Dona M. Wong, *The Wall Street journal guide to information graphics: the dos and don'ts of presenting data, facts, and figures*, 1st ed. New York: W.W. Norton & Co, 2010, ISBN: 978-0-393-07295-2.

Slide 29

References

[Cleveland 1989] William S. Cleveland, *The elements of graphing data*, 10.[print.] ed. Monterey, Cal: Wadsworth, 1989, ISBN: 978-0-534-03729-1.

[Cleveland 1993] William S. Cleveland, *Visualizing data*. Murray Hill, N.J. : [Summit, N.J: At&T Bell Laboratories ; Published by Hobart Press, 1993, ISBN: 978-0-9634884-0-4.

[Lyons 2010] Ray Lyons, 'Best Practices in Graphical Data Presentation', Baltimore, MD, USA, 25-Oct-2010 [Online]. Available: http://libraryassessment.org/bm~doc/workshop_lyons_ray.pdf. [Accessed: 18-Aug-2015]

[Tufte 2008] Edward Rolf Tufte, *Envisioning information*, 12. printing. Cheshire, Conn: Graphics Press, 2008, ISBN: 978-0-9613921-1-6.

[Tufte 1997] Edward R. Tufte, *Visual explanations: images and quantities, evidence and narrative*. Cheshire, Conn: Graphics Press, 1997, ISBN: 978-0-9613921-2-3.

[Tufte 2006] Edward R. Tufte, *Beautiful evidence*. Cheshire, Conn: Graphics Press, 2006, ISBN: 978-0-9613921-7-8.

[Tufte 2001] Edward R. Tufte, *The visual display of quantitative information*, 2nd ed. Cheshire, Conn: Graphics Press, 2001, ISBN: 978-0-9613921-4-7.

[Wong 2010] Dona M. Wong, *The Wall Street journal guide to information graphics: the dos and don'ts of presenting data, facts, and figures*, 1st ed. New York: W.W. Norton & Co, 2010, ISBN: 978-0-393-07295-2.

Slide 30