Un hackathon dataviz pour aider une ONG

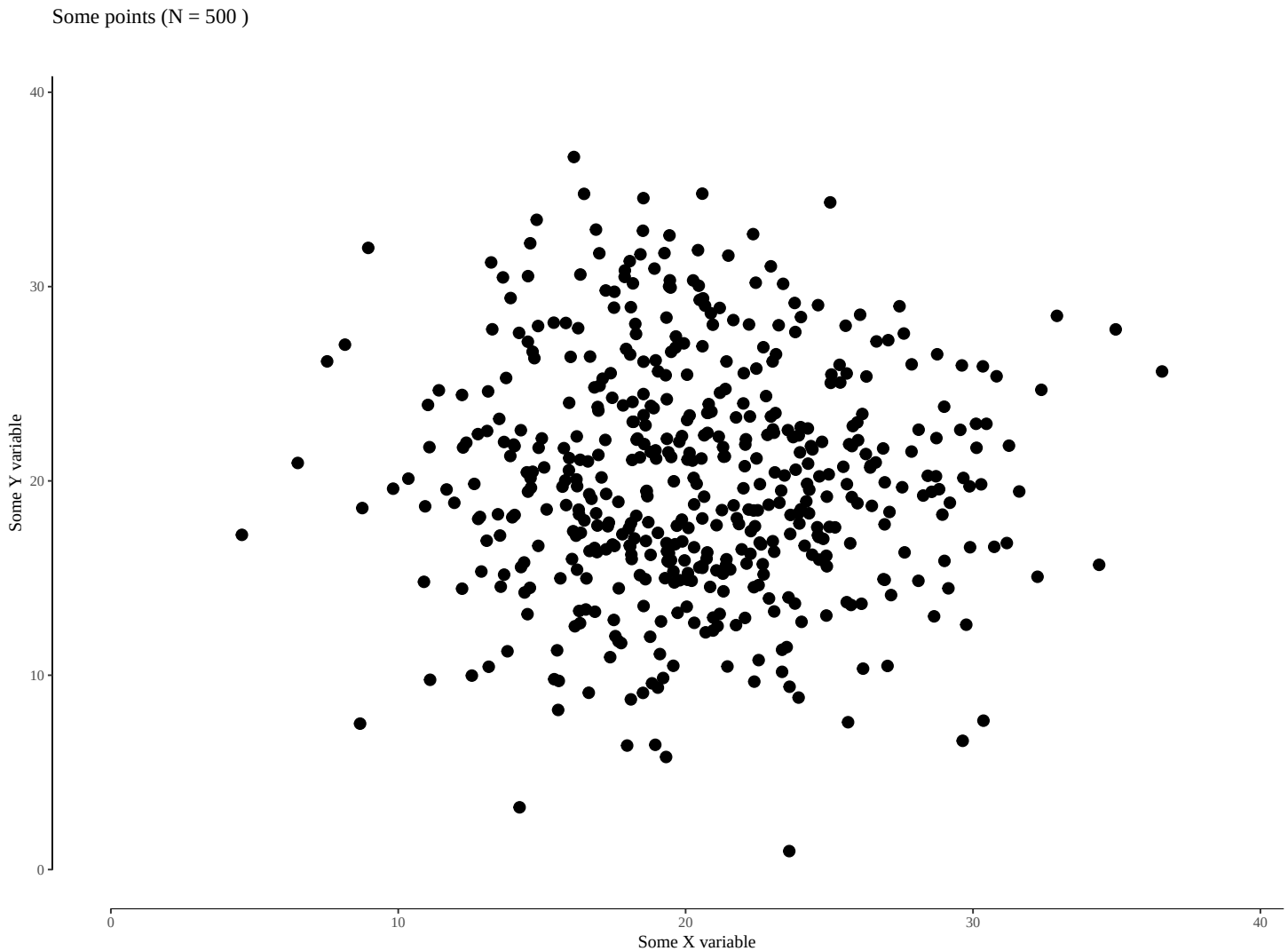
De la théorie aux (bonnes) pratiques de la data visualisation

Mars, 2019

Christophe Bontemps
Toulouse School of Economics (INRA)
&
Édith Maulandi
Viz For SocialGood

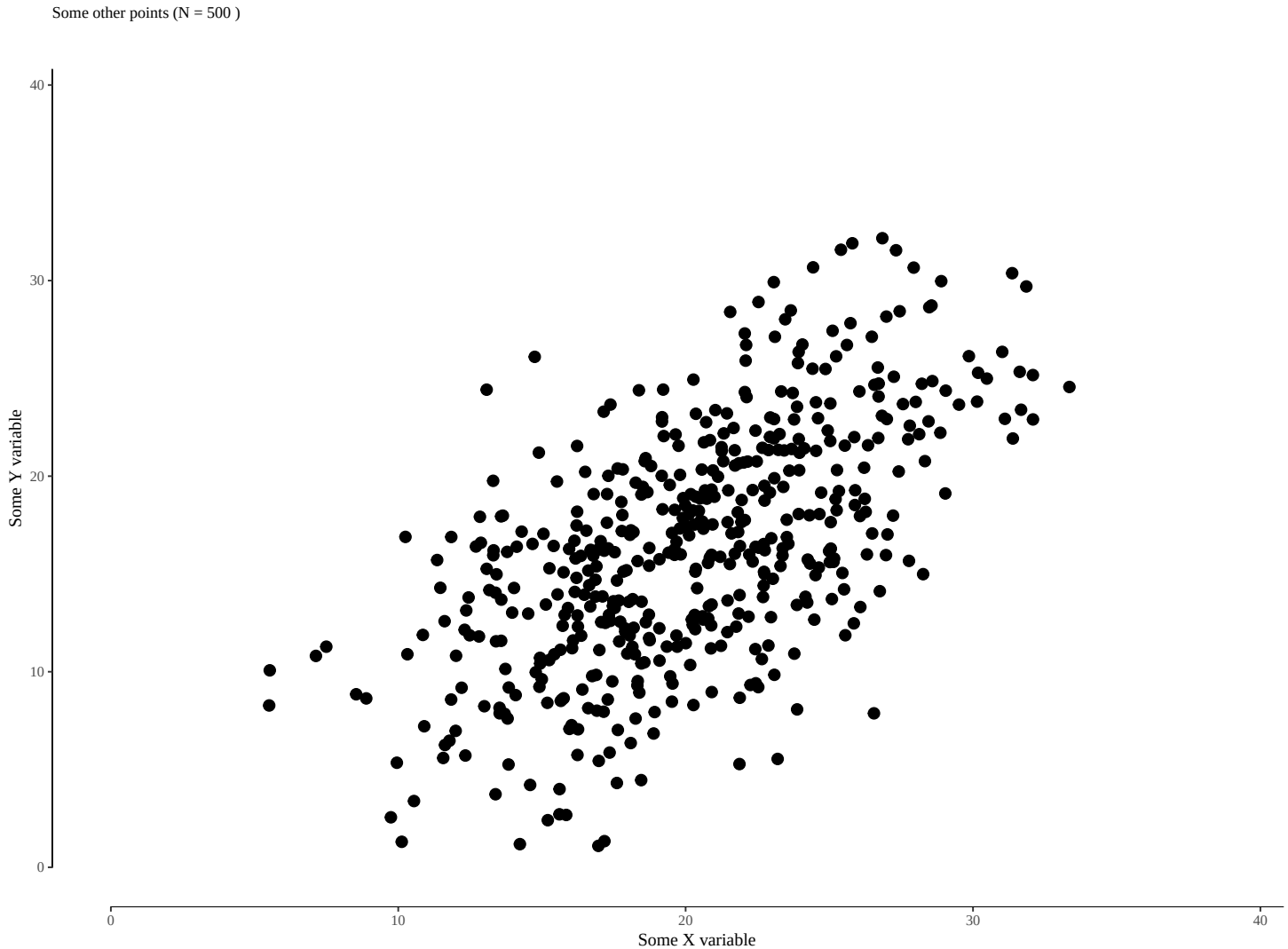
THE "VISUAL PERCEPTION" OF A GRAPHIC

What do you see ?

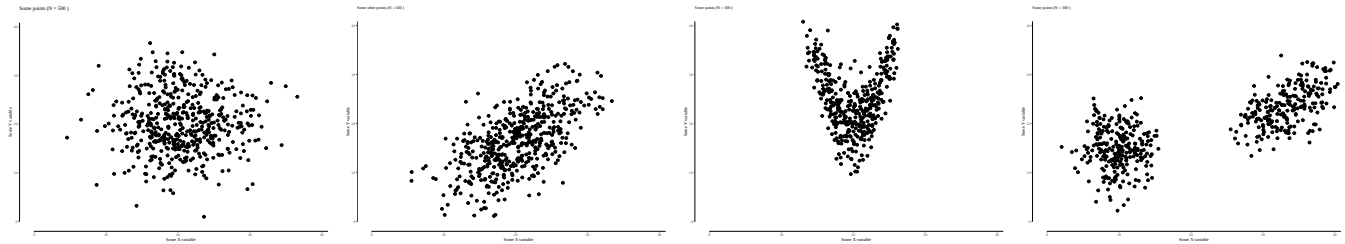


THE "VISUAL PERCEPTION" OF A GRAPHIC

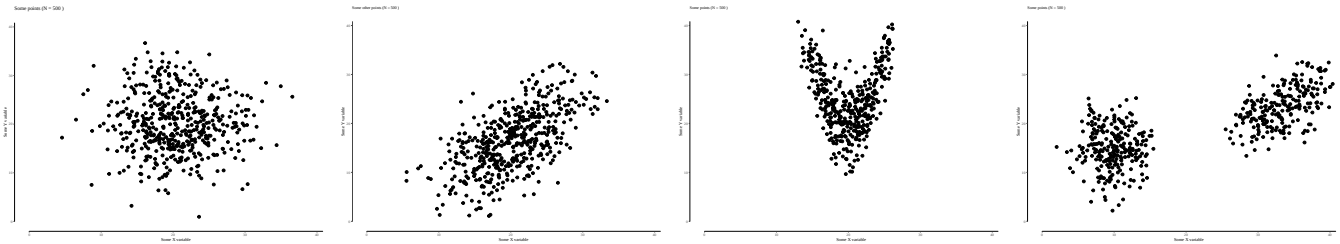
And here, what do you see ?



"Visual perception" as a statistical test

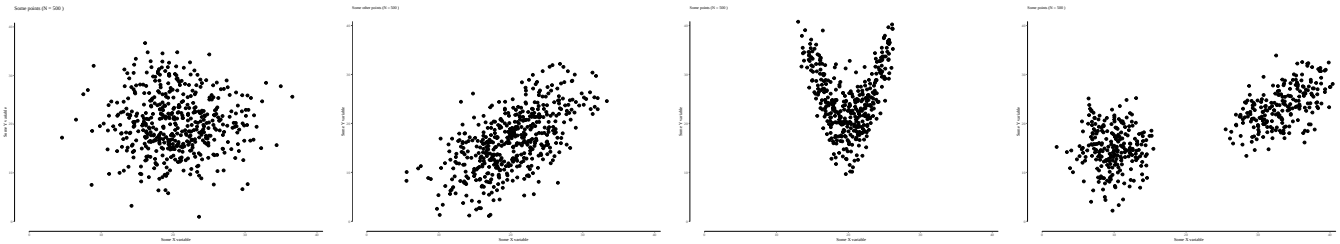


"Visual perception" as a statistical test



" The human eye acts is a broad feature detector and general statistical test". Buja et al. (2009)

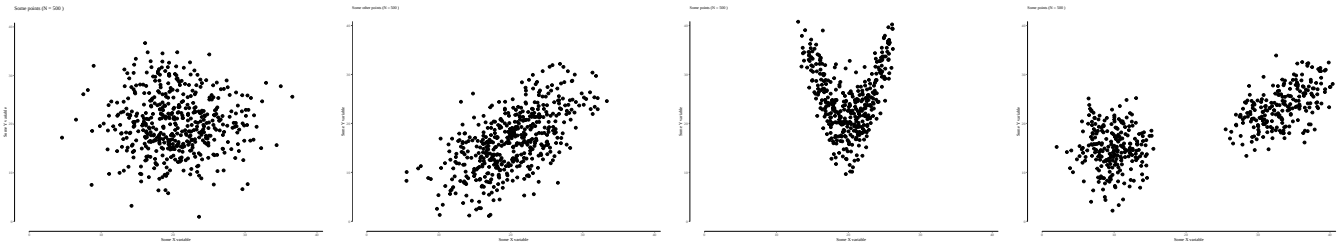
"Visual perception" as a statistical test



" The human eye acts is a broad feature detector and general statistical test". Buja et al. (2009)

Test : H_0 : {There is "nothing" } = {**No** relation}

"Visual perception" as a statistical test

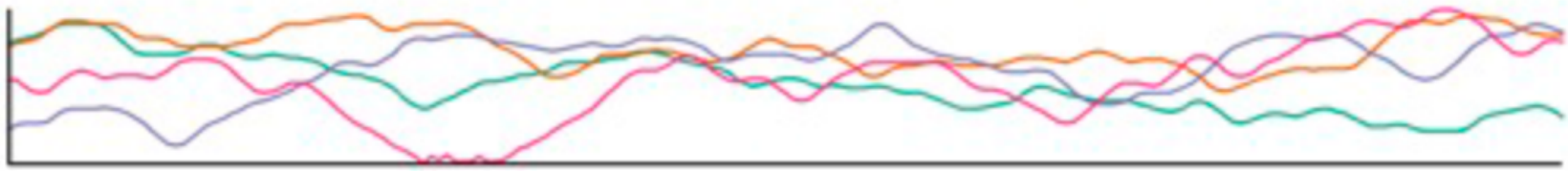


" The human eye acts is a broad feature detector and general statistical test". Buja et al. (2009)

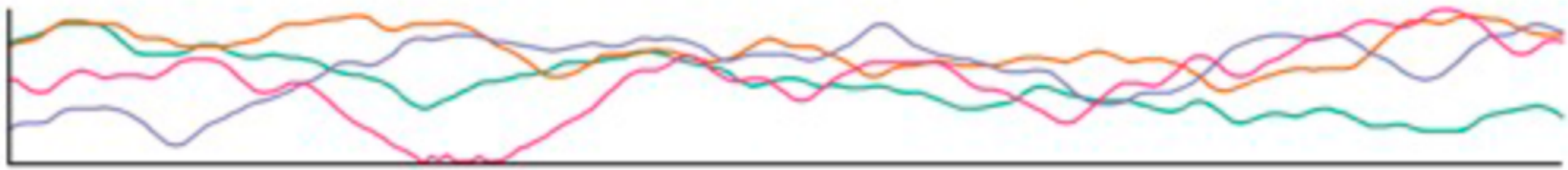
Test : H_0 : {There is "nothing" } = {**No** relation}

H_1 : { There is "something" } = {There is **some** relation (Correlation, linearity, heterogeneity, groups..) }

"Visual perception" as a comparison

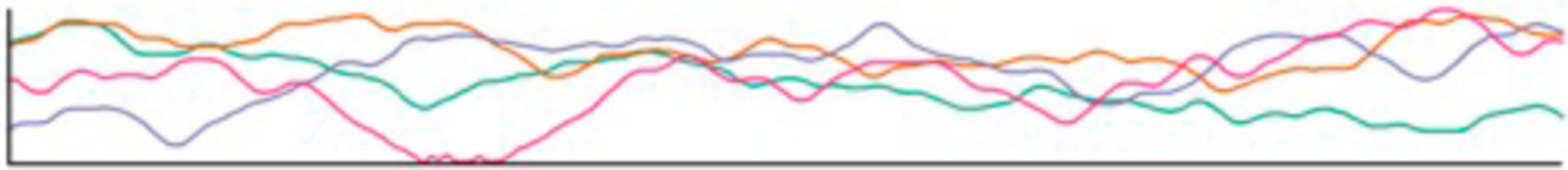


"Visual perception" as a comparison



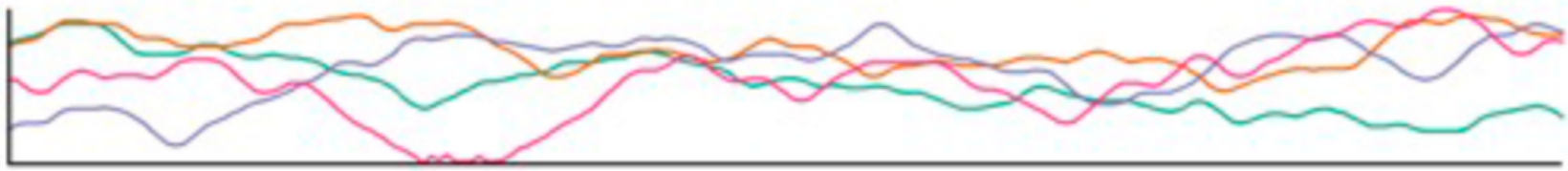
- What do you see here ?

"Visual perception" as a comparison



- What do you see here ?

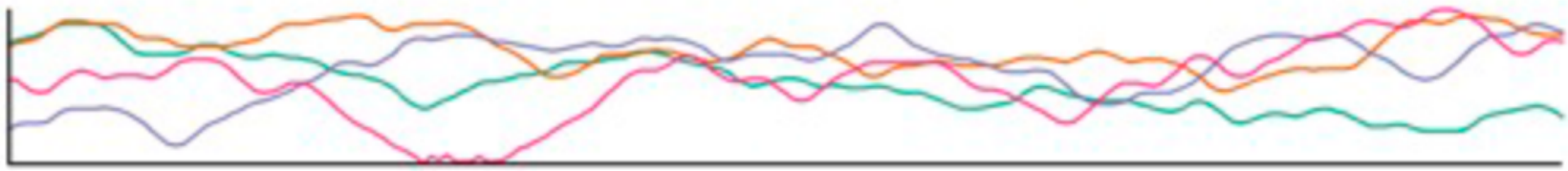
"Visual perception" as a comparison



- What do you see here ?

Difficult to see the maximum/minimum of each curve...

"VISUAL PERCEPTION" AS A COMPARISON



- What do you see here ?

Difficult to see the maximum/minimum of each curve...

Idea shared by Gelman (2004) and Munzner (2014)

ROAD MAP : **A REVOIR**

- What is data visualisation ?
  - What for ? What questions ?
  - Type of graphics, classics, Case studies
  - Does dynamic help ?
- Rules
  - Bertin, Tufte and Cleveland's rules
  - Tables *vs* graphics
  - Visual perception
- Visualizing complexity
  - Mixing variables types
  - Visualising many
  - *Maps and networks*
  - *Case studies*
- *Interactions*

Goals of Data Visualisation

Data visualisation serves at least two main purposes

- Data exploration
  Graphs as visual tests, comparisons → **short** time to built and to read

Goals of Data Visualisation

Data visualisation serves at least two main purposes

- Data exploration
  Graphs as visual tests, comparisons → **short** time to built and to read
- Data representation
  Summaries, comparisons, storytelling → **long** time to build, short time to read

Goals of Data Visualisation

Data visualisation serves at least two main purposes

- Data exploration
  Graphs as visual tests, comparisons → **short** time to built and to read
- Data representation
  Summaries, comparisons, storytelling → **long** time to build, short time to read

  The problem is that :

> *" Communicating implies **simplification***
> *data exploration implies **exhaustivity**"*

THE BIG PICTURE : FROM "DATA TO VIZ"

The Big Picture : from "data to viz"

THE BIG PICTURE : FROM "DATA TO VIZ"

bone_fsort	bx2_fsort	rmsefsort	bone_ftrue	bx2_ftrue	rmseftrue	bone_probit	bx2_probit
.8867406	1.070447	3.590005	.7361284	1.400824	3.134265	.8609696	1.231284
.8081753	1.043795	3.152994	.8169689	1.354571	5.354777	1.100465	.7198337
1.078269	1.019714	4.50032	1.002494	1.070324	3.478615	1.063318	1.009623
.9313936	.7030163	3.239488	.9822327	.6482554	2.657284	.7893744	1.192139
.9105234	.9432839	3.073645	1.128867	.1960131	3.048621	1.021027	1.051435
.9291375	.9827668	5.349204	.9653457	.9511749	3.337612	1.037403	.8830764
1.113444	.6736442	3.027501	1.074889	.8294352	2.768067	1.020905	1.165862
.8570632	1.468859	4.519791	.848403	1.732238	4.459501	.9893017	1.27312
.6581307	1.612068	5.042707	.4977604	2.136239	8.321213	1.058817	.6954145
1.012545	.8085507	3.354877	1.070338	.6093824	2.682671	1.080817	.722274
1.32485	.5022374	4.440361	1.124536	.6874371	2.85944	1.015451	1.096031
.8012164	1.282709	4.629036	.7592914	1.696804	5.805454	.8978575	1.240651
1.293513	.4332167	7.53362	1.139851	.6672146	3.874777	1.051809	.878014
.8227926	1.325787	3.790141	.8705111	1.189695	3.497889	1.050775	1.014878
1.131029	.4216127	3.397707	1.197935	.5421426	3.498553	1.031771	.907118
.8022098	1.176128	3.059879	.9859373	1.07031	3.116614	1.044137	1.013132
.9142329	1.031845	3.334867	1.095129	.8095155	2.97967	1.106252	.8586126
.8031349	1.414903	6.319849	1.118786	.7290513	2.998966	1.130253	.9202099
1.189141	.714771	4.009688	.9306182	1.310915	4.476381	.9666906	1.08258
.4775752	1.956697	4.588708	.4204233	2.137922	5.132593	.8733798	1.252393
1.096664	.9689751	7.59197	1.070467	.70661	4.728706	1.067702	.8673179
.7572182	1.476068	4.22933	.9634842	.9691185	2.875876	1.00177	.8843074
.9754681	.9435486	3.809161	1.035667	.6965184	2.457987	1.147595	.9096023
.9490274	.778796	3.063304	.5720246	1.253536	9.415662	.9240479	1.133063
1.038485	.7553858	3.330338	.9900171	.9530999	3.031227	1.042053	1.035369
.9724179	1.435002	5.77195	.9292588	1.36202	3.811148	1.062936	1.037366
1.04662	.8743635	4.122163	1.04075	.9089242	3.230211	.8933463	1.326798
.9540951	1.254786	5.074101	.9994549	1.089691	4.616098	1.02264	.9741149
.9190806	.9136872	3.734955	1.020362	.8567674	3.020249	.913802	1.144934
.9661756	.7474627	3.238612	.96892	.7516584	2.464803	.9460078	.9226835
1.022502	.8197547	3.618911	1.115418	.6282563	3.218334	1.180864	.6482964
.9044181	.906302	3.507935	.8862579	.9701268	2.548325	.8544677	1.201299
1.103058	.4806759	4.708026	1.002306	.6740937	2.896619	1.015808	.8446978
.7439321	1.282491	3.520073	.820206	1.081162	2.788524	.8678131	1.258859

THE BIG PICTURE : FROM "DATA TO VIZ"

Coding

bone_fsort	bx2_fsort	rmsefsort	bone_ftrue	bx2_ftrue	rmseftrue	bone_probit	bx2_probit
.8867406	1.070447	3.590005	.7361284	1.400824	3.134265	.8609696	1.231284
.8081753	1.043795	3.152994	.8169689	1.354571	5.354777	1.100465	.7198337
1.078269	1.019714	4.50032	1.002494	1.070324	3.478615	1.063318	1.009623
.9313936	.7030163	3.239488	.9822327	.6482554	2.657284	.7893744	1.192139
.9105234	.9432839	3.073645	1.128867	.1960131	3.048621	1.021027	1.051435
.9291375	.9827668	5.349204	.9653457	.9511749	3.337612	1.037403	.8830764
1.113444	.6736442	3.027501	1.074889	.8294352	2.768067	1.020905	1.165862
.8570632	1.468859	4.519791	.848403	1.732238	4.459501	.9893017	1.27312
.6581307	1.612068	5.042707	.4977604	2.136239	8.321213	1.058817	.6954145
1.012545	.8085507	3.354877	1.070338	.6093824	2.682671	1.080817	.722274
1.32485	.5022374	4.440361	1.124536	.6874371	2.85944	1.015451	1.096031
.8012164	1.282709	4.629036	.7592914	1.696804	5.805454	.8978575	1.240651
1.293513	.4332167	7.53362	1.139851	.6672146	3.874777	1.051809	.878014
.8227926	1.325787	3.790141	.8705111	1.189695	3.497889	1.050775	1.014878
1.131029	.4256127	3.397707	1.197935	.5421426	3.498553	1.031771	.907118
.8022098	1.176128	3.059879	.9859373	1.07031	3.116614	1.044137	1.013132
.9142329	1.031845	3.334867	1.095129	.6095151	2.97967	1.106252	.8586126
.8031349	1.414903	6.319849	1.118786	.7290513	2.998966	1.130253	.9202099
1.189141	.714771	4.009688	.9306182	1.310915	4.476381	.9666906	1.08258
.4775752	1.956697	4.588708	.4204233	2.137922	5.132593	.8733798	1.252393
1.096664	.9689751	7.59197	1.070467	.70661	4.728706	1.067702	.8673179
.7572182	1.476068	4.22933	.9634842	.9691185	2.875876	1.00177	.8843074
.9754681	.9435486	3.809161	1.035667	.6965184	2.457987	1.147595	.9096023
.9490274	.778796	3.063304	.5720246	1.253536	9.415662	.9240479	1.133063
1.038485	.7553858	3.330338	.9900171	.9583099	3.031227	1.042053	1.035369
.9724179	1.435002	5.77195	.9292588	1.36202	3.811148	1.062936	1.037366
1.04662	.8743635	4.122163	1.04075	.9089242	3.230211	.8933463	1.326798
.9540951	1.254786	5.074101	.9994549	1.089691	4.616098	1.02264	.9741149
.9190806	.9136872	3.734955	1.020362	.8567674	3.020249	.913802	1.144934
.9661756	.7474627	3.238612	.96892	.7516584	2.464803	.9460078	.9226835
1.022502	.8197547	3.618911	1.115418	.6282563	3.218334	1.180864	.6482964
.9044181	.906302	3.507935	.8862579	.9701268	2.548325	.8544677	1.201299
1.103058	.4806759	4.708026	1.002306	.6740937	2.896619	1.015808	.8446978
.7439321	1.282491	3.520073	.820206	1.081162	2.788524	.8678131	1.258859

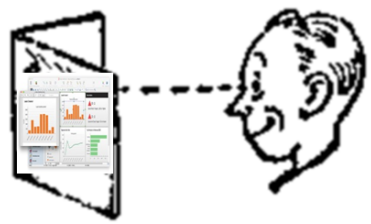
The Big Picture : from "data to viz"

Coding

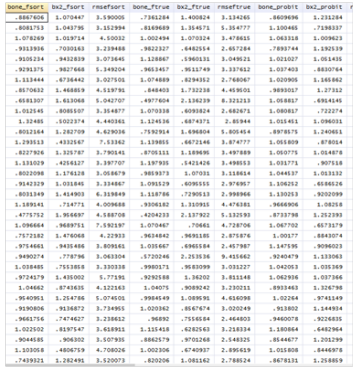
bone_fsort	bx2_fsort	rmsefsort	bone_ftrue	bx2_ftrue	rmseftrue	bone_probit	bx2_probit
.8867606	1.070447	3.590005	.7361284	1.440824	3.134265	-.8609696	1.231284
.8081753	1.043795	3.152994	.8169689	1.354571	5.354777	1.100465	.7198337
1.078269	1.019714	4.50032	1.002494	1.070324	3.478615	1.063318	1.009623
.9313936	.7030163	3.239488	.9822327	.6482554	2.657284	.7893744	1.192139
.9105234	.9432839	3.073645	1.128867	.1960131	3.048621	1.021027	1.051435
.9291375	.9827668	5.349204	.9653457	.9511749	3.337612	1.037403	.8830764
1.113444	.6736442	3.027501	1.074889	.8294352	2.768067	1.020905	1.165862
.8570632	1.468859	4.519791	.848403	1.732238	4.459501	.9893017	1.27312
.6581307	1.612068	5.042707	.4977604	2.136239	8.321213	1.058817	.6954145
1.012545	.8085507	3.354877	1.070338	.6093824	2.682671	1.080817	.722274
1.32485	.5022374	4.440361	1.124536	.6874371	2.85944	1.015451	1.096031
.8012164	1.282709	4.629036	.7592914	1.696804	5.805454	.8978575	1.240651
1.293513	.4332167	7.53362	1.139851	.6672146	3.874777	1.051809	.878014
.8227926	1.325787	3.790141	.8705111	1.189695	3.497889	1.050775	1.014878
1.131029	.4216127	3.397707	1.197935	.5421426	3.498553	1.031771	.907118
.8022098	1.176128	3.059879	.9859373	1.07031	3.116614	1.044137	1.013132
.9142329	1.031845	3.334867	1.095129	.6095155	2.97967	1.106252	.6586126
.8031349	1.414903	6.319849	1.118786	.7290513	2.998966	1.130253	.9202099
1.189141	.714771	4.009688	.9306182	1.310915	4.476381	.9666906	1.08258
.4775752	1.956697	4.568708	.4204233	2.137922	5.132593	.8733798	1.252393
1.096664	.9689751	7.59197	1.070467	.70661	4.728706	1.067702	.6873179
.7572182	1.476068	4.22933	.9634842	.9691185	2.875876	1.00177	.8843074
.9754661	.9435486	3.809161	1.035667	.6965184	2.457987	1.147595	.9096023
.9490274	.778796	3.063304	.5720248	1.253536	9.415662	.9240479	1.133063
1.038485	.7553858	3.330338	.9900171	.9583099	3.031227	1.042053	1.035369
.9724179	1.435002	5.77195	.9292588	1.36202	3.811148	1.062936	1.037366
1.04662	.8743635	4.122163	1.04075	.9089242	3.230211	.8933463	1.326798
.9540951	1.254786	5.074101	.9994549	1.089691	4.616098	1.02264	.9741149
.9190806	.9136872	3.734955	1.020362	.8567674	3.020249	.913802	1.144934
.9661756	.7474627	3.238612	.96892	.7516584	2.464803	.9460078	.9226835
1.022502	.8197547	3.618911	1.115418	.6282563	3.218334	1.180864	.6482964
.9044185	.906302	3.507935	.8862579	.9701268	2.548325	.8544677	1.201299
1.103058	.4806759	4.708026	1.002306	.6740937	2.896619	1.015808	.8446978
.7439321	1.282491	3.520073	.820206	1.081162	2.788524	.8678131	1.258859

THE BIG PICTURE : FROM "VIZ TO DATA"

Decoding ?



THE BIG PICTURE : FROM "VIZ TO DATA"

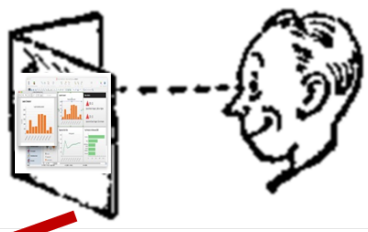


bone_fsort	bx2_fsort	rmsefsort	bone_ftrue	bx2_ftrue	rmseftrue	bone_probit	bx2_probit
.8867406	1.070447	3.590005	.7361284	1.400824	3.134265	-.8609696	1.231284
.8081753	1.043795	3.152994	.8169689	1.354571	5.354777	1.100465	.7198337
1.078269	1.019714	4.50032	1.002494	1.070324	3.478615	1.063318	1.009623
.9313936	.7030183	3.239488	.9822327	.6482554	2.657284	.7893744	1.192539
.9105234	.9432839	3.073646	1.128867	.1960131	3.048621	1.021027	1.011435
.9291375	.9827668	5.349204	.9653457	.9511749	3.337612	1.037403	.8830764
1.113444	.6736442	3.027501	1.074889	.8294352	2.768067	1.020905	1.165862
.8570632	1.468859	4.519791	.848403	1.732238	4.459501	.9893017	1.27312
.6581307	1.612068	5.042707	.4977604	2.136239	8.321213	1.058817	.6954145
1.012545	.8085507	3.354877	1.070338	.6093824	2.682671	1.080817	.722274
1.32485	.5022374	4.440361	1.124536	.6874371	2.85944	1.015451	1.096031
.8012164	1.282709	4.629036	.7592914	1.696804	5.805454	.8978575	1.240651
1.293513	.4332167	7.53362	1.139851	.6672146	3.874777	1.051809	.878014
.8227926	1.325787	3.790141	.8705111	1.189696	3.497889	1.050775	1.014878
1.131029	.4216127	3.397707	1.197935	.5421426	3.498553	1.031771	.907518
.8022098	1.176128	3.059879	.9819373	1.07031	3.116614	1.044137	1.013132
.9142329	1.031845	3.334867	1.095129	.6095155	2.974967	1.106252	.6586126
.8031349	1.414903	6.319849	1.118786	.7290513	2.998966	1.130253	.9202099
1.189141	.714771	4.009688	.9306182	1.310915	4.476381	.9666906	1.08258
.4775752	1.956697	4.588708	.4204233	2.137922	5.132593	.8733798	1.252393
1.096664	.9689751	7.592197	1.070467	.70661	4.728706	1.067702	.6873379
.7572182	1.476068	4.22933	.9634842	.9691185	2.875876	1.00177	.8843074
.9754681	.9435486	3.809161	1.035667	.6965184	2.457987	1.147595	.9096023
.9490274	.778796	3.063304	.5720246	1.253536	9.415662	.9240479	1.133063
1.038485	.7553858	3.330338	.9900171	.9530999	3.033227	1.042053	1.035369
.9724179	1.435002	5.77195	.9292588	1.36202	3.811148	1.062936	1.037366
1.04662	.8743635	4.122163	1.04075	.9089242	3.230211	.8933463	1.326798
.9540951	1.254786	5.074101	.9994549	1.089691	4.616098	1.02264	.9741149
.9190806	.9136872	3.734955	1.020362	.8567674	3.020249	.913802	1.144934
.9661756	.7474627	3.238612	.96892	.7516584	2.464803	.9460078	.9226835
1.022502	.8197547	3.618911	1.115418	.6282563	3.218334	1.180864	.6482964
.9044181	.906302	3.507935	.8862579	.9701268	2.548325	.8544677	1.201299
1.103058	.4806759	4.708026	1.002306	.6740937	2.896619	1.015808	.8446978
.7439321	1.282491	3.520073	.820206	1.081162	2.788524	.8678131	1.258859

Decoding

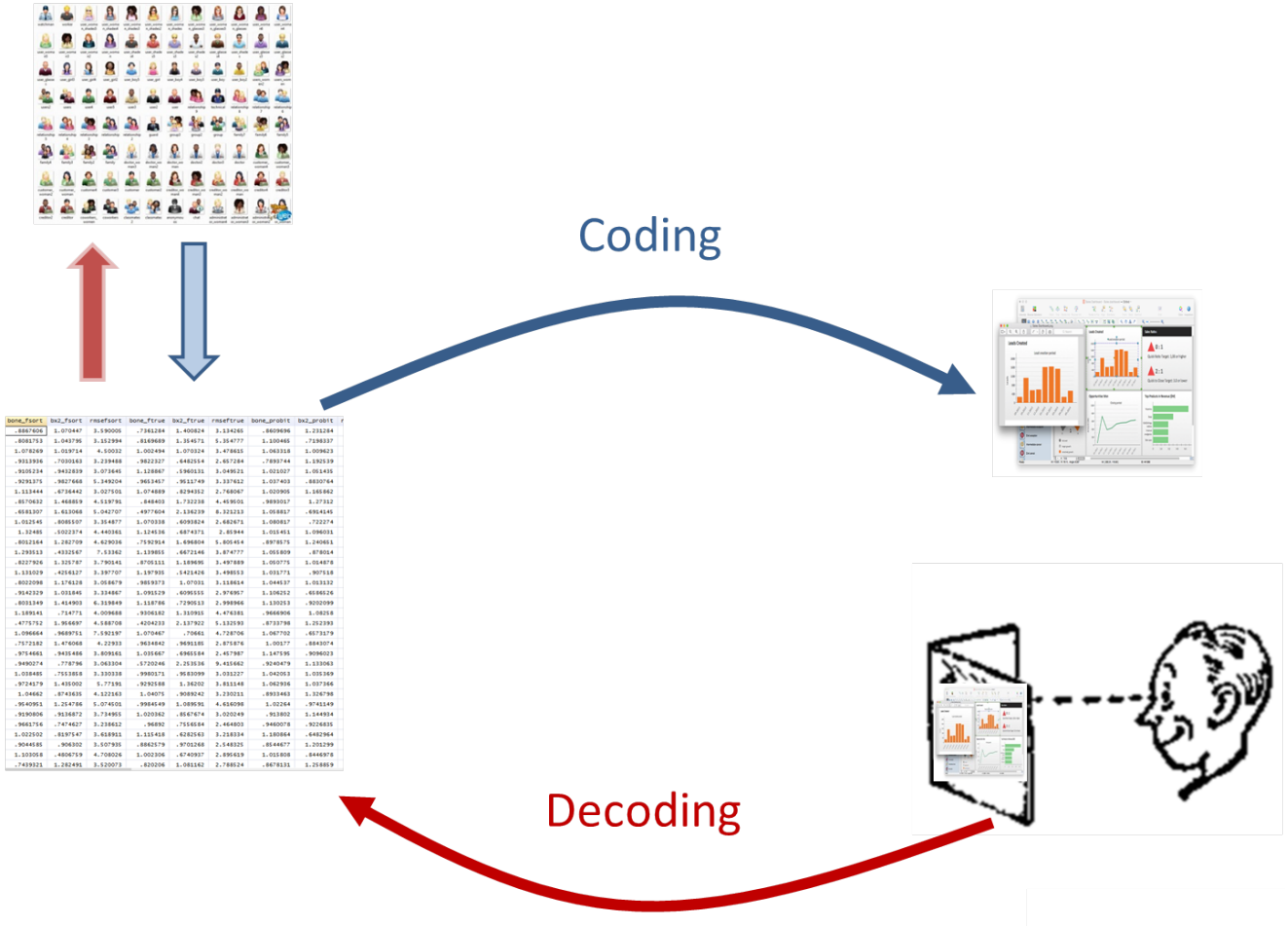
THE BIG PICTURE : FROM "VIZ TO DATA"

bone_fsort	bx2_fsort	rmsefsort	bone_ftrue	bx2_ftrue	rmseftrue	bone_probit	bx2_probit
.8867406	1.070447	3.590005	.7361284	1.400824	3.134265	.8609696	1.231284
.8081753	1.043795	3.152994	.8149689	1.354571	5.354777	1.100465	.7198337
1.078269	1.019714	4.50032	1.002494	1.070324	3.478615	1.063318	1.009623
.9313936	.7030163	3.239488	.9822327	.6482554	2.657284	.7893744	1.192539
.9105234	.9432839	3.073646	1.128867	.1960131	3.048621	1.021027	1.011435
.9291375	.9827668	5.349204	.9653457	.9511749	3.337612	1.037403	.8830764
1.113444	.6736442	3.027501	1.074889	.8294352	2.768067	1.020905	1.165862
.8570632	1.468859	4.519791	.848403	1.732238	4.459501	.9893017	1.27312
.6581307	1.612068	5.042707	.4977604	2.136239	8.321213	1.058817	.6954145
1.012545	.8085507	3.354877	1.070338	.6093824	2.682671	1.080817	.722274
1.32485	.5022374	4.440361	1.124536	.6874371	2.85944	1.011451	1.096031
.8012164	1.282709	4.629036	.7592914	1.696804	5.805454	.8978575	1.240651
1.293513	.4332167	7.53362	1.139851	.6672146	3.874777	1.051809	.878014
.8227926	1.325787	3.790141	.8705111	1.189696	3.497889	1.050775	1.014878
1.131029	.4216127	3.397707	1.197935	.5421426	3.498553	1.031771	.907118
.8022098	1.176128	3.059879	.9819373	1.07031	3.116614	1.044137	1.013132
.9142329	1.031845	3.334867	1.091529	.6095151	2.974967	1.106252	.6586126
.8031349	1.414903	6.319849	1.118786	.7290513	2.998966	1.130253	.9202099
1.189141	.714771	4.009688	.9306182	1.310915	4.476381	.9666906	1.08258
.4775752	1.956697	4.568708	.4204233	2.137922	5.132593	.8733798	1.252393
1.096664	.9689751	7.592197	1.070467	.70661	4.728706	1.067702	.8673379
.7572182	1.476068	4.22933	.9634842	.9691185	2.875876	1.00177	.8843074
.9754681	.9435486	3.809161	1.035667	.6965184	2.457987	1.147595	.9096023
.9490274	.778796	3.063304	.5720246	1.253536	9.415662	.9240479	1.133063
1.038485	.7553858	3.330338	.9900171	.9530999	3.033227	1.042053	1.035369
.9724179	1.435002	5.77191	.9292588	1.36202	3.811148	1.062936	1.037366
1.04662	.8743635	4.122163	1.04075	.9089242	3.230211	.8933463	1.326798
.9540951	1.254786	5.074101	.9994549	1.089691	4.616098	1.02264	.9741149
.9190806	.9136872	3.734955	1.020362	.8567674	3.020249	.913802	1.144934
.9661756	.7474627	3.238612	.96892	.7516584	2.464803	.9460078	.9226835
1.022502	.8197547	3.618911	1.115418	.6282563	3.218334	1.180864	.6482964
.9044181	.906302	3.507935	.8862579	.9701268	2.548325	.8544677	1.201299
1.103058	.4806759	4.708026	1.002306	.6740937	2.896619	1.015808	.8446978
.7439321	1.282491	3.520073	.820206	1.081162	2.788524	.8678131	1.258859

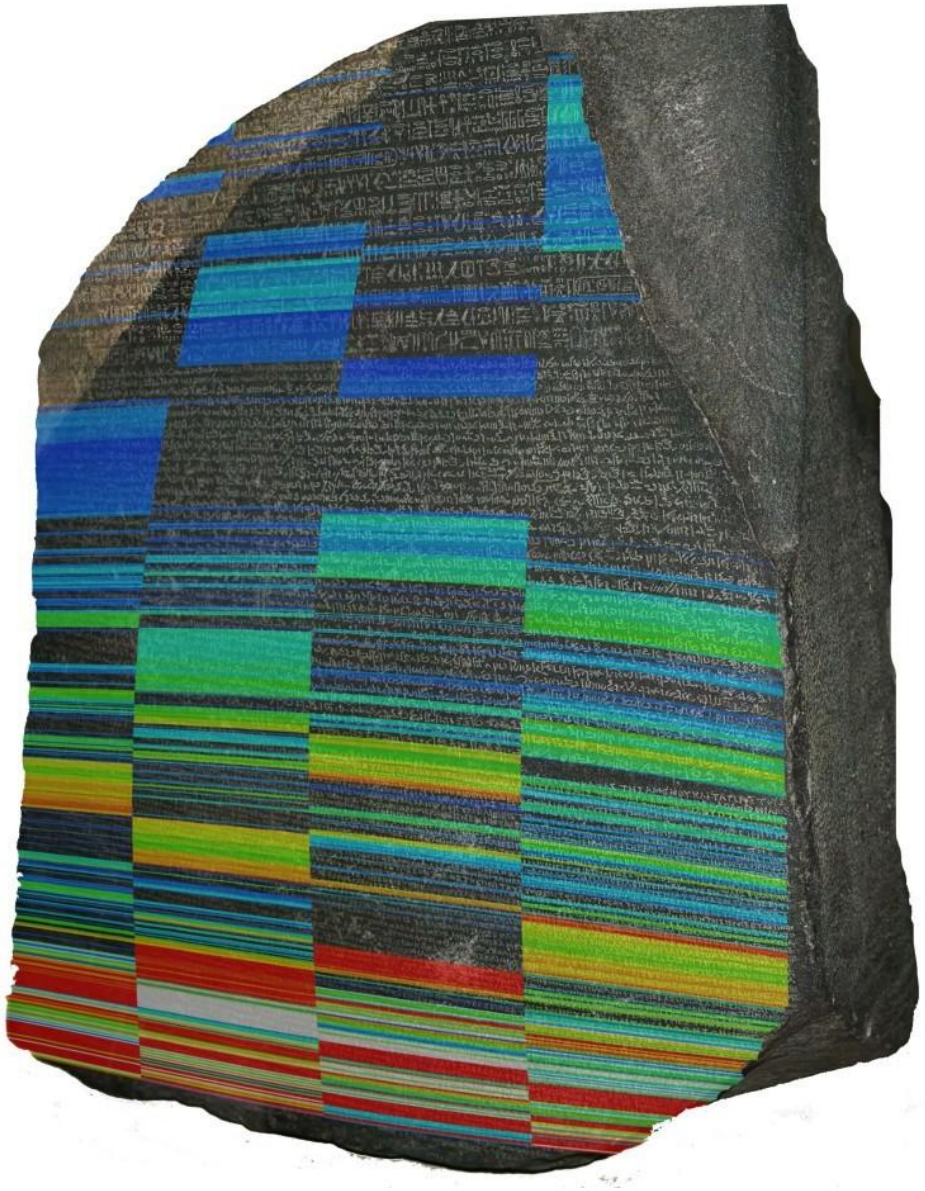


Decoding

THE BIG PICTURE : FROM "VIZ TO DATA"



[- Decoding visual information -]



GRAPHICS IMMEDIATE TO UNDERSTAND

FIGURE – Where do people run in Paris

source : (N. Yau)
http://flowingdata.com/2014/02/05/where-people-run/

GRAPHICS IMMEDIATE TO UNDERSTAND

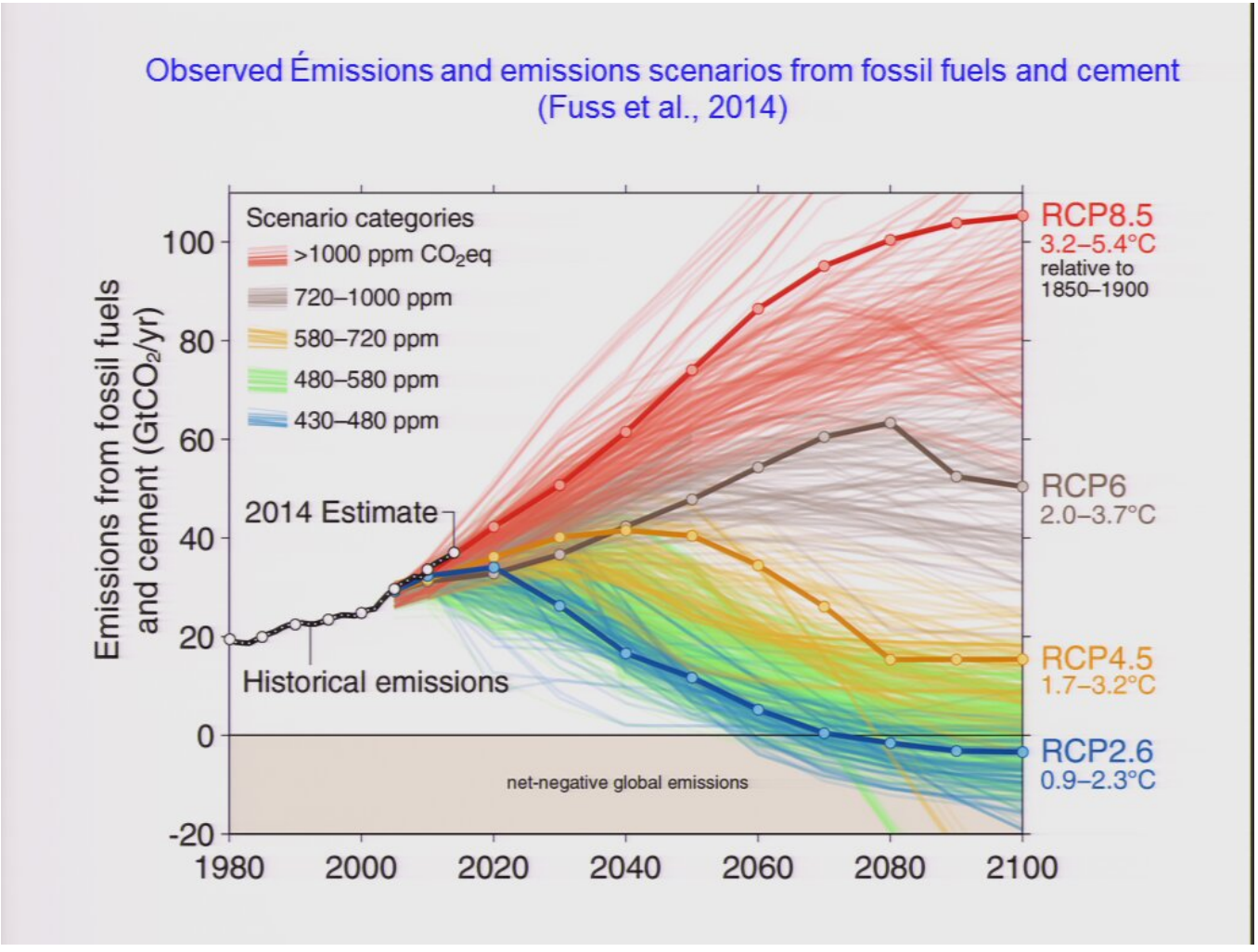


FIGURE – Climate forecast uncertainty (S. Planton)

GRAPHICS THAT NEED EXPLANATIONS :

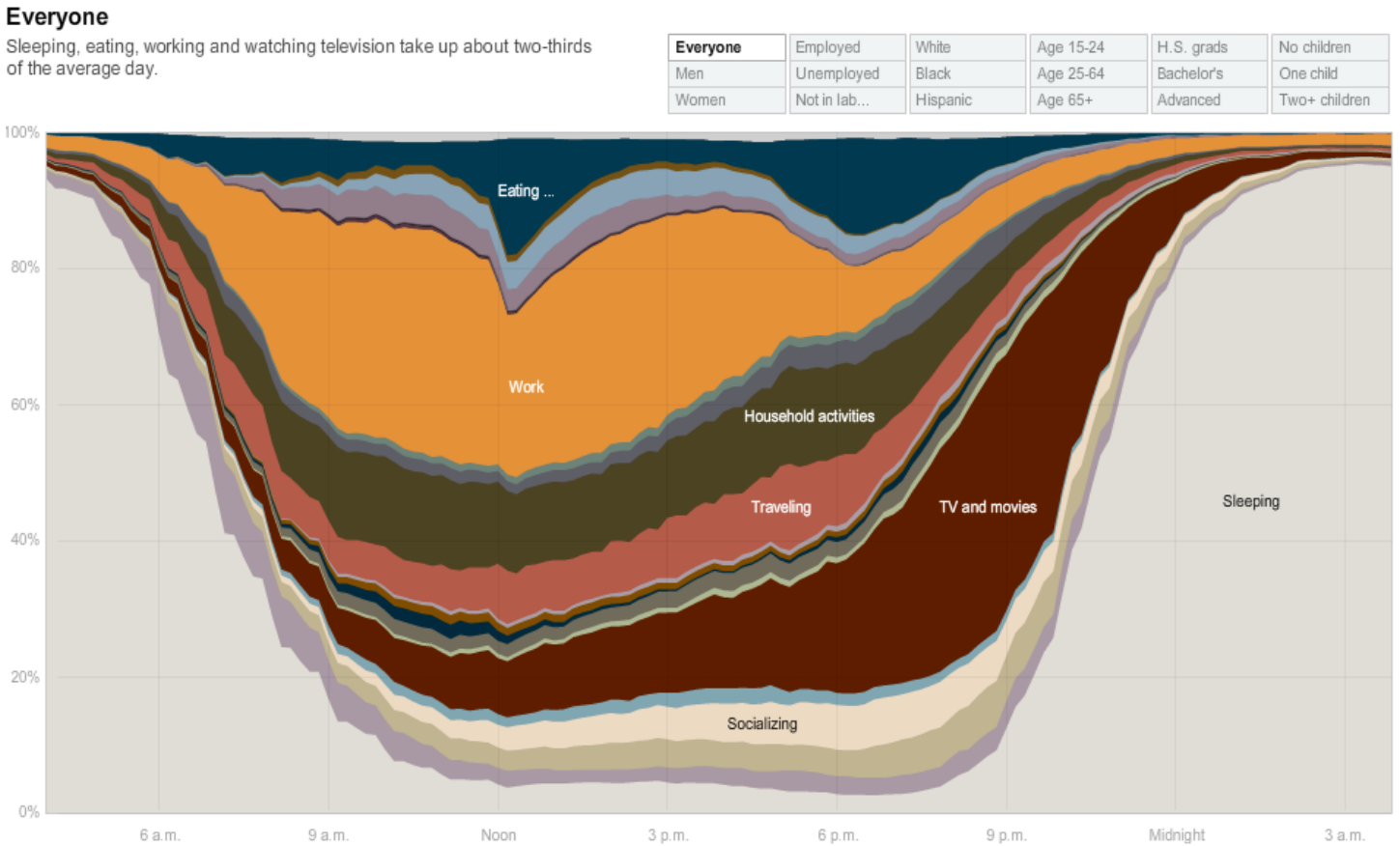


FIGURE – How people spend their days (NYT).

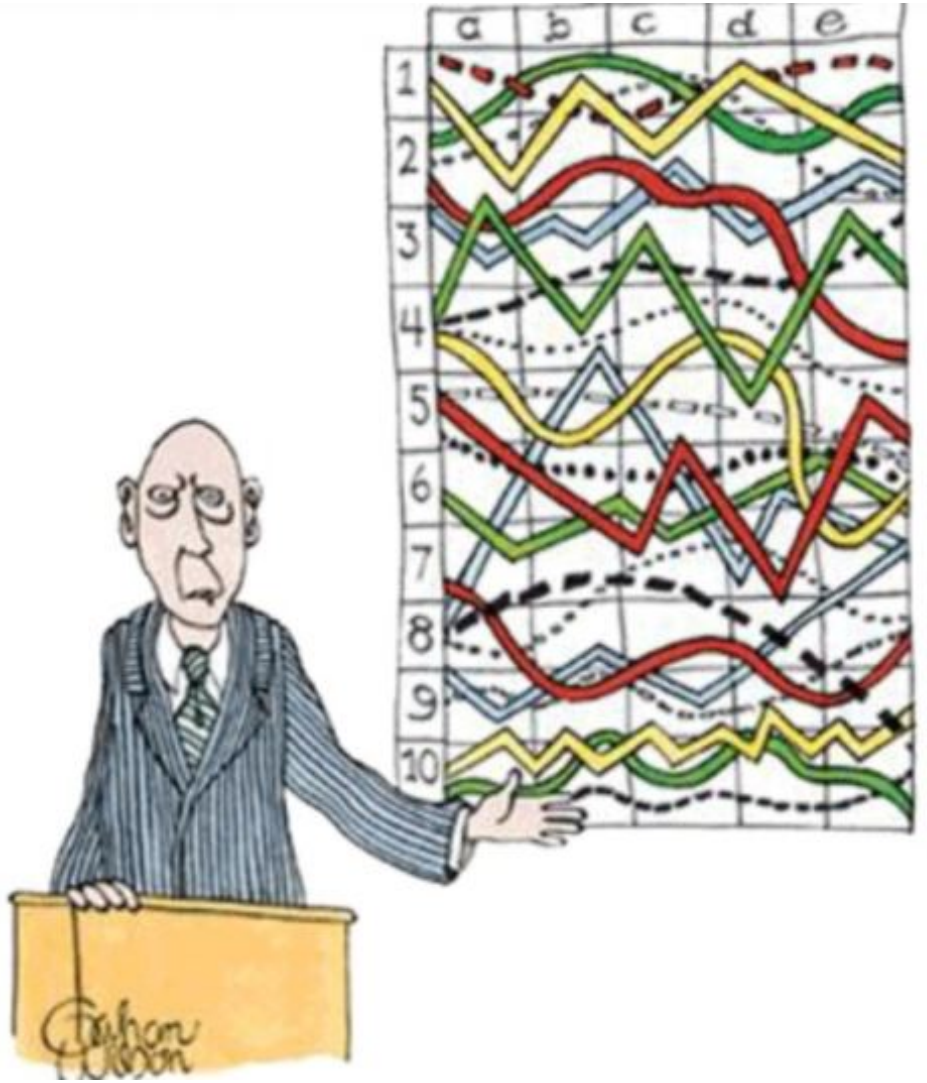
So what are the rules ?

- Can you name some rules for a good graphic ?

So what are the rules ?

- Can you name some rules for a good graphic ?
- Or maybe, rules for a "bad" graphic first..

[- MISTAKES ! -]



"I'll pause for a moment so you can let this information sink in."

Source : New Yorker

Usual Mistakes

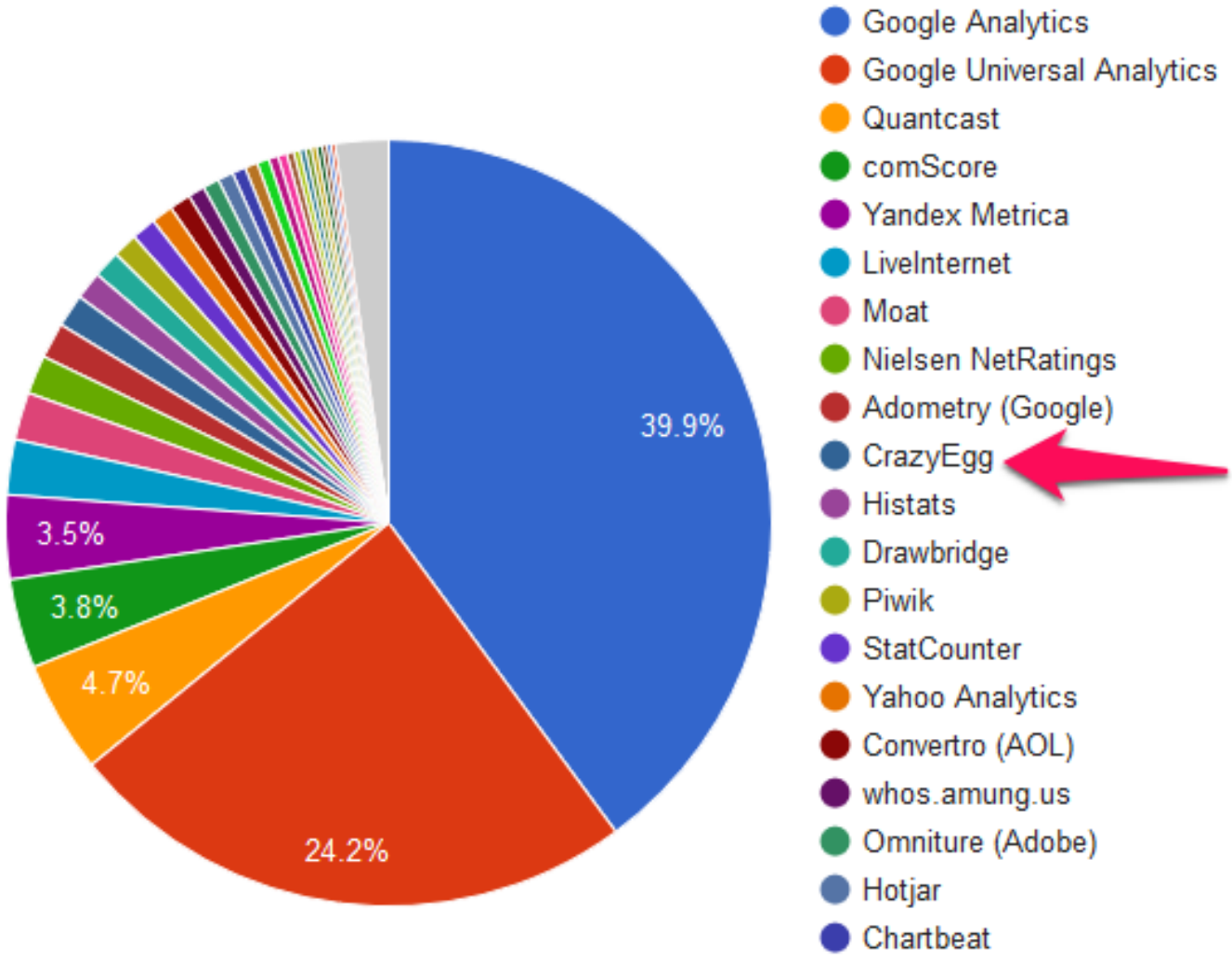


Figure – (source WTF Visualization)

EVEN WORSE : 3D-PIECHART

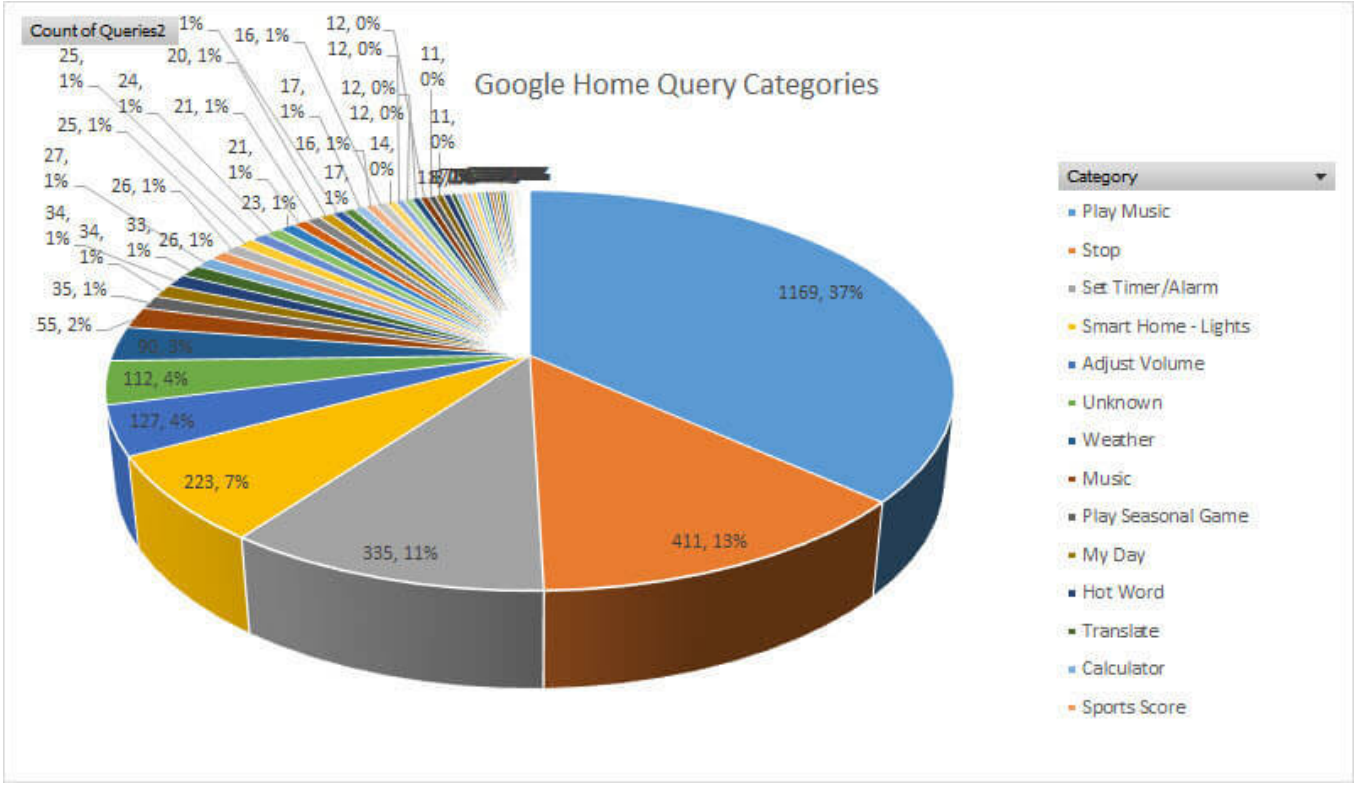


FIGURE – (source WTF Visualization)

SOME PEOPLE USE IT !

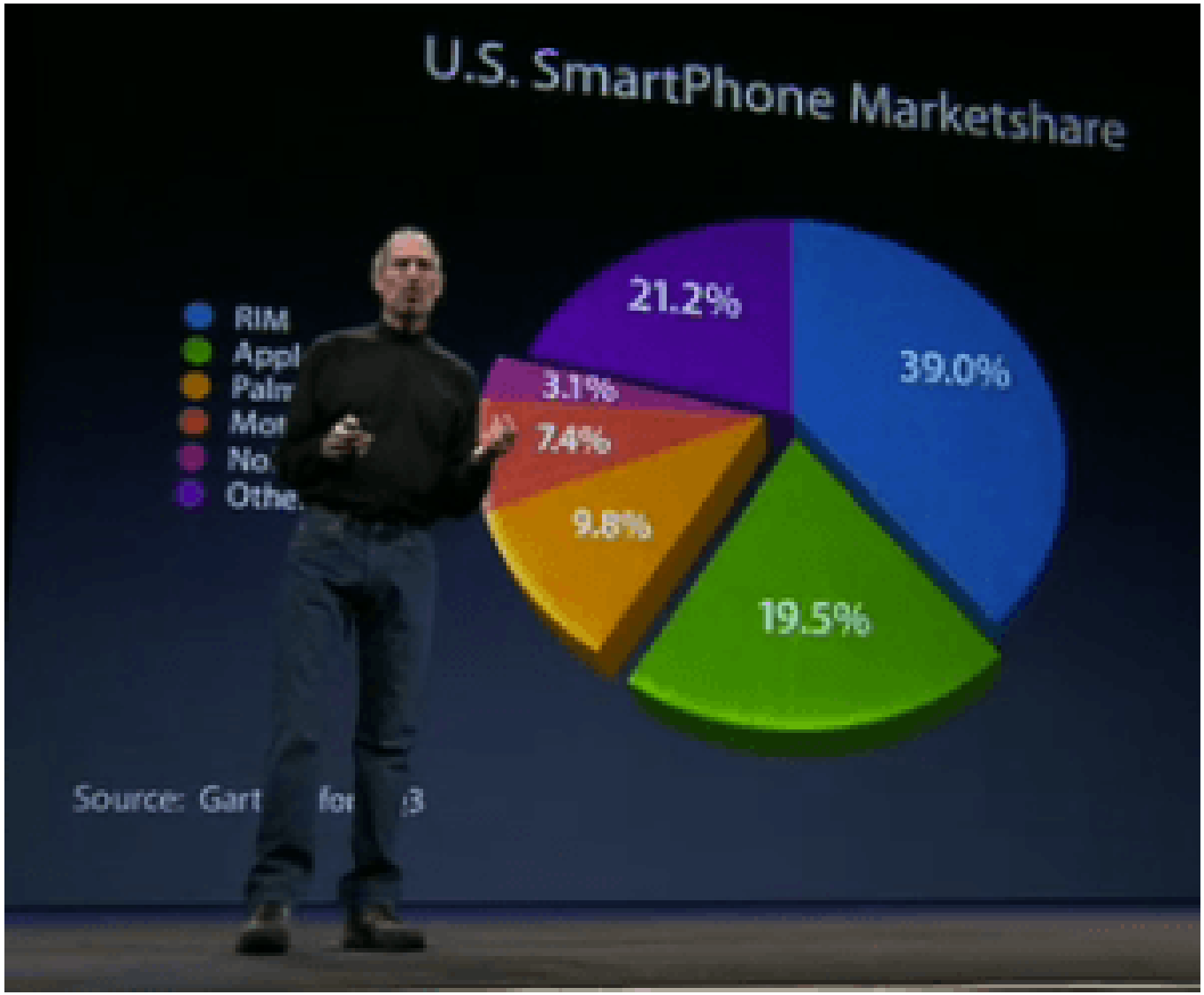


FIGURE – Source : 5 mistakes in Dataviz

WHY IS IT A BAD PRACTICE ?

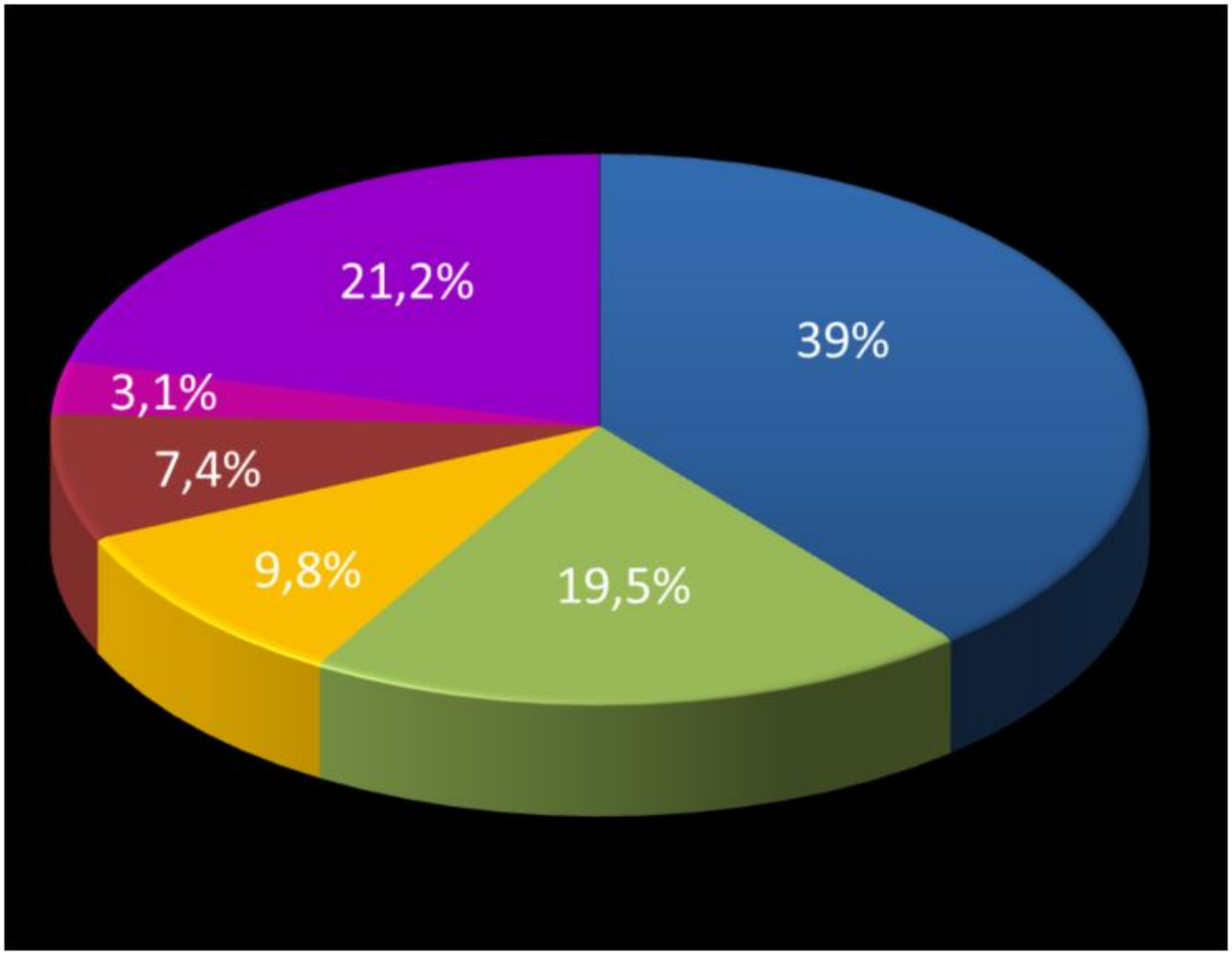


FIGURE – Source : reproducing Steve Job's dataviz (C. Bontemps)

THE AREAS DO NOT REPRESENT THE DATA

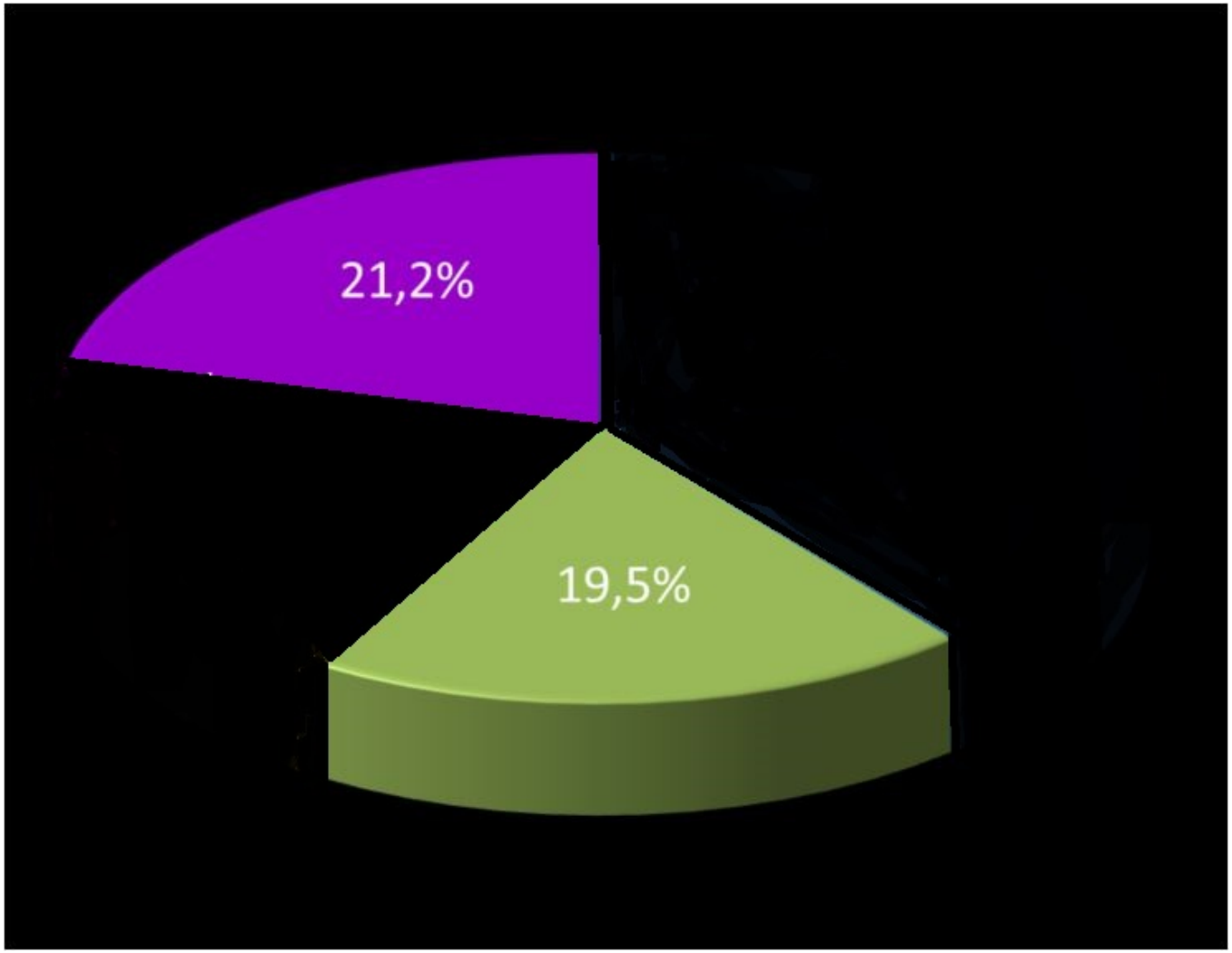


FIGURE – Source : reproducing Steve Job's dataviz (C. Bontemps)

TOP : $20cm^2 = 21.2\%$, BOTTOM : $30\ cm^2 = 19.5\%$

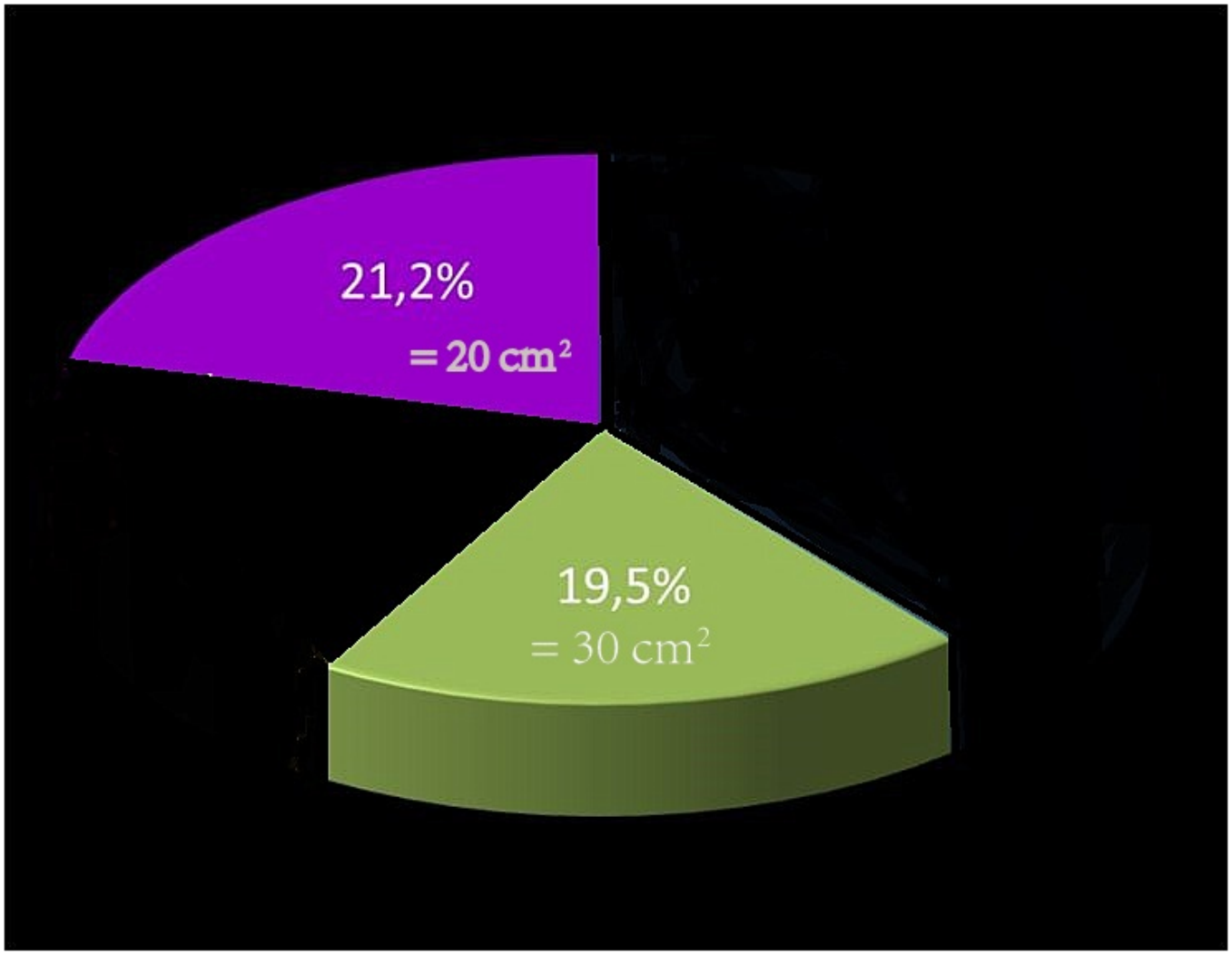


FIGURE – Source : reproducing Steve Job's dataviz (C. Bontemps)

USUAL MISTAKES

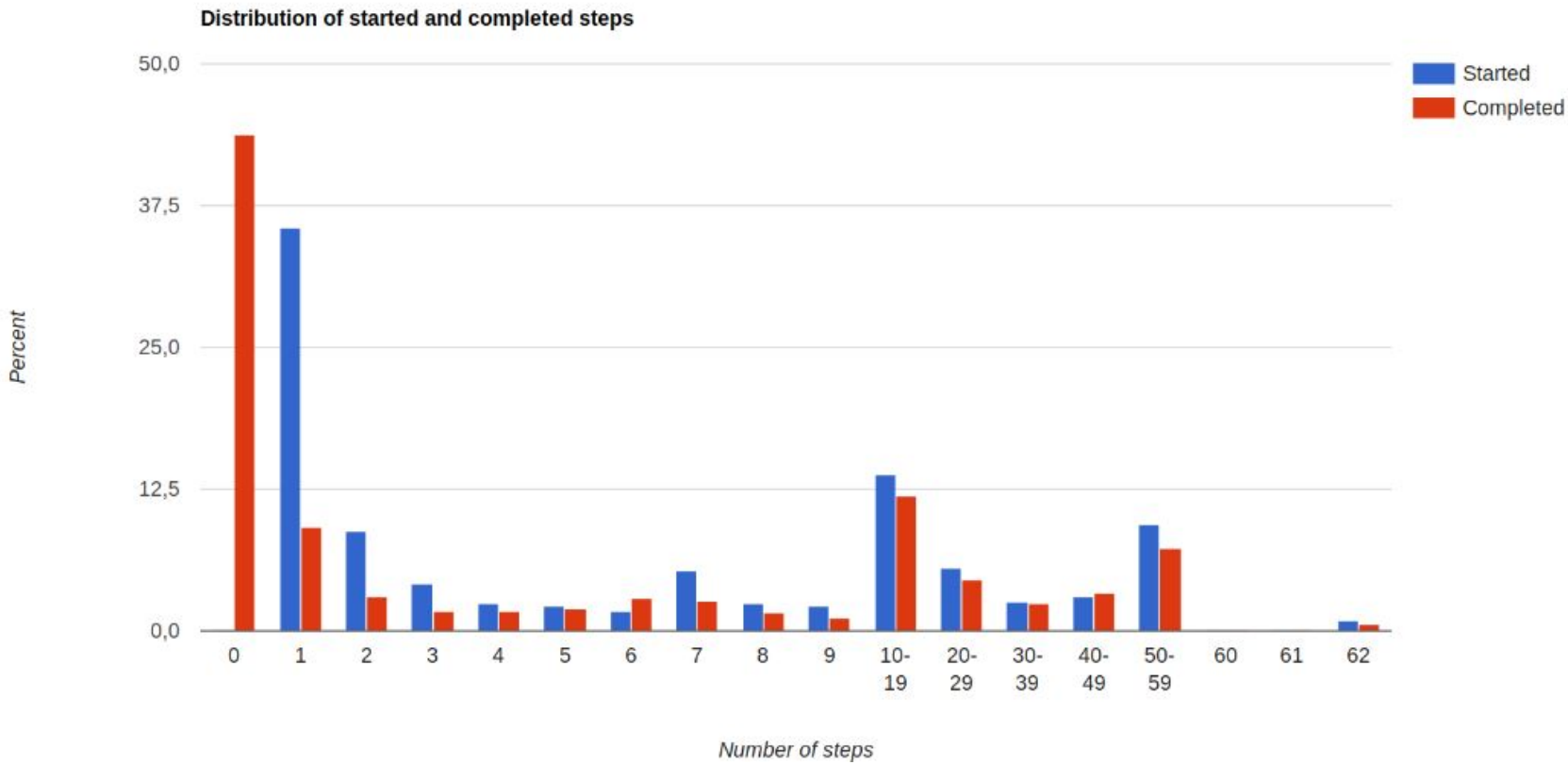


FIGURE – Percentage of student starting and completing a "step" in MOOC

Source : My 2017' students (Jan & Mohamed), but also recent researchers' presentations

WHAT'S WRONG WITH THIS GRAPHIC ?

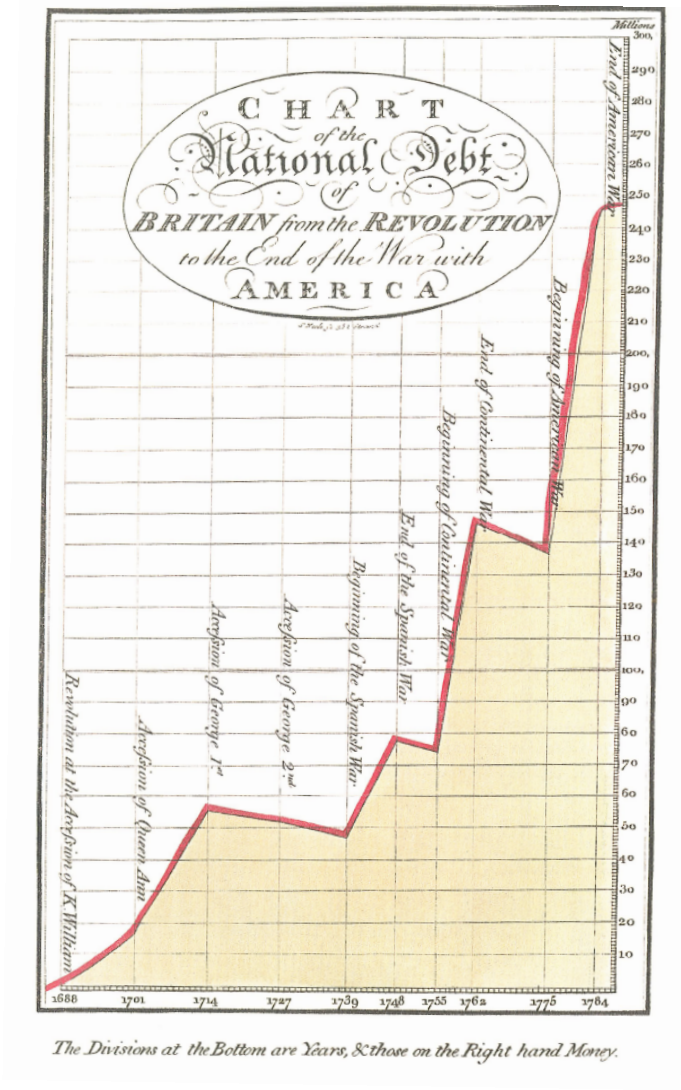


FIGURE – Government spending "Skyrocketing". Source : Tufte (2001) from Playfair(1786)

SCALES ARE MISLEADING !

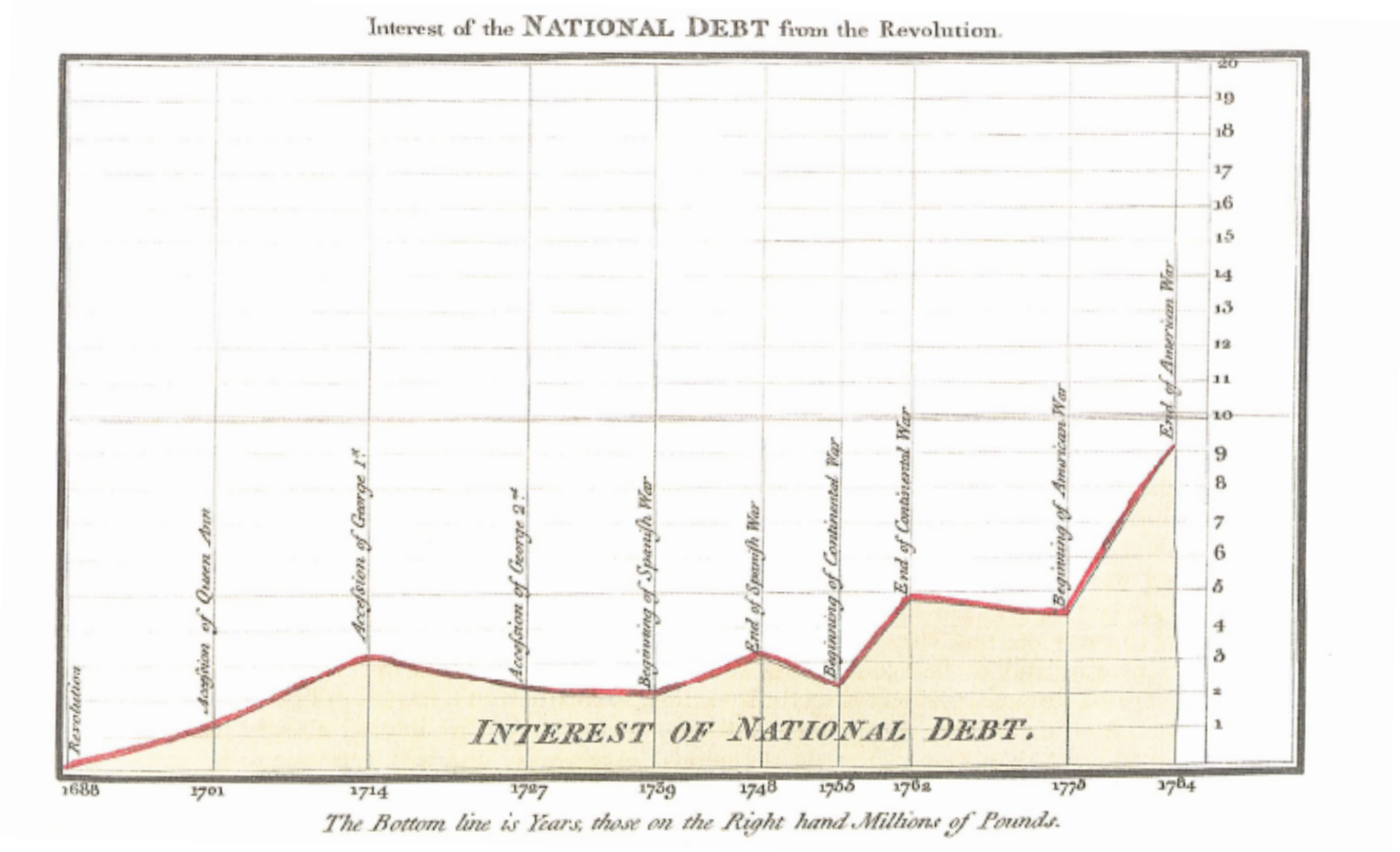


FIGURE – Governemnt spending "Skyrocketing" (revisited)- Source : Tufte (2001) from Playfair(1786).

NOWADAYS TOO :

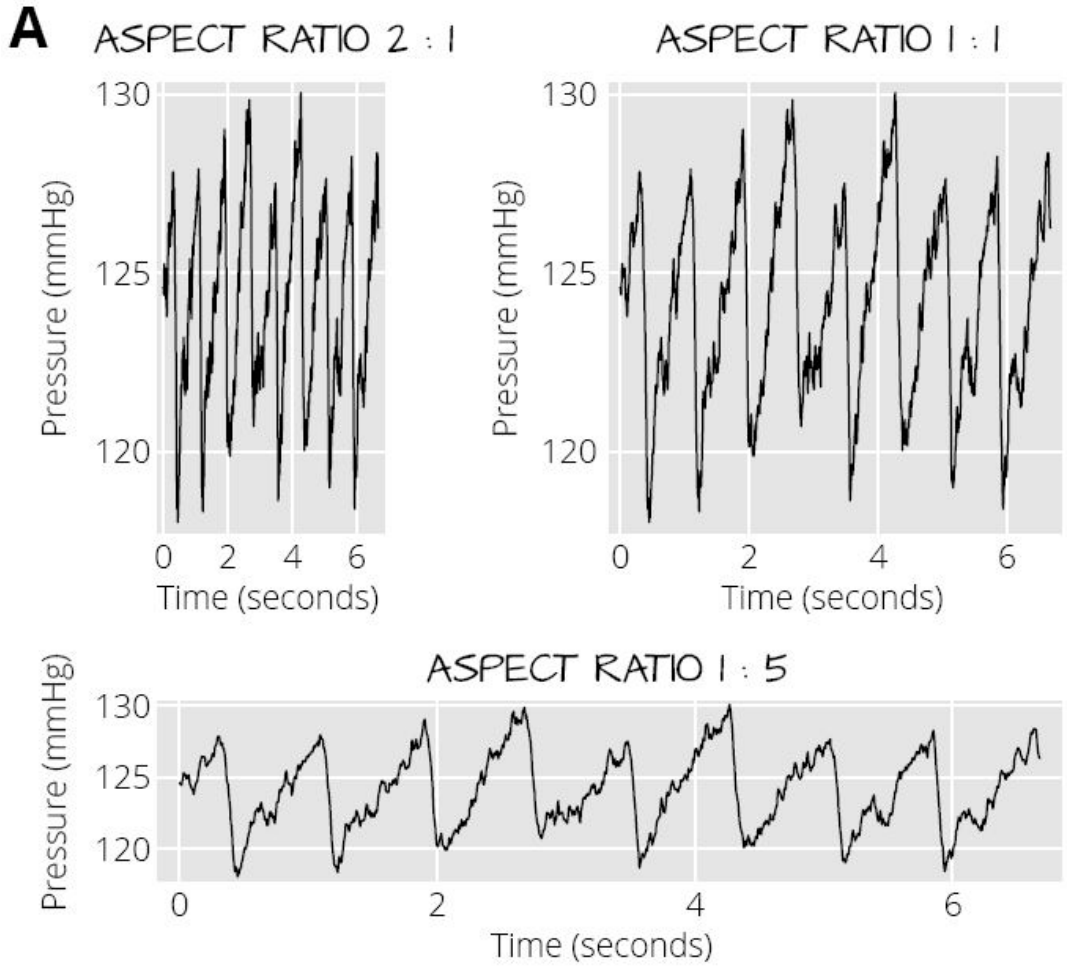


FIGURE – From Allen and Erhardt (2016b)

NOWADAYS TOO :

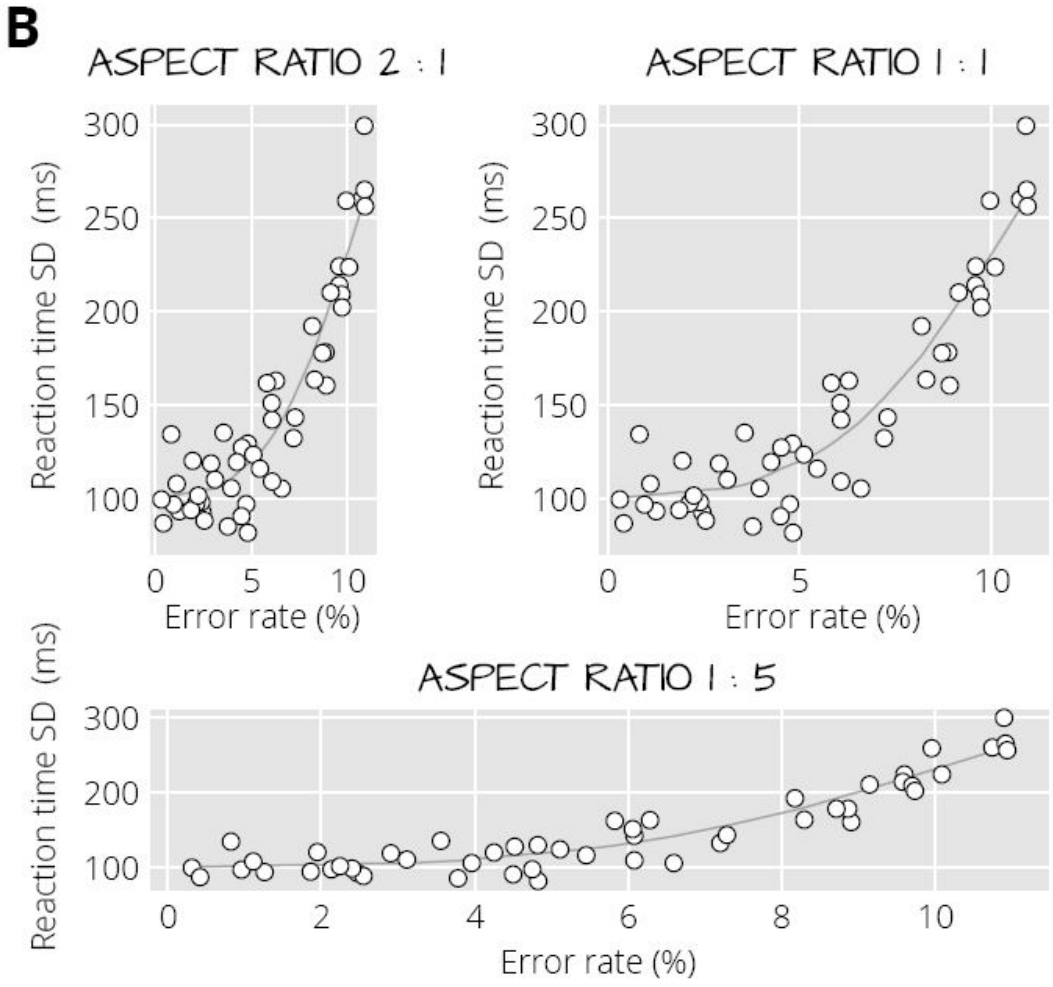
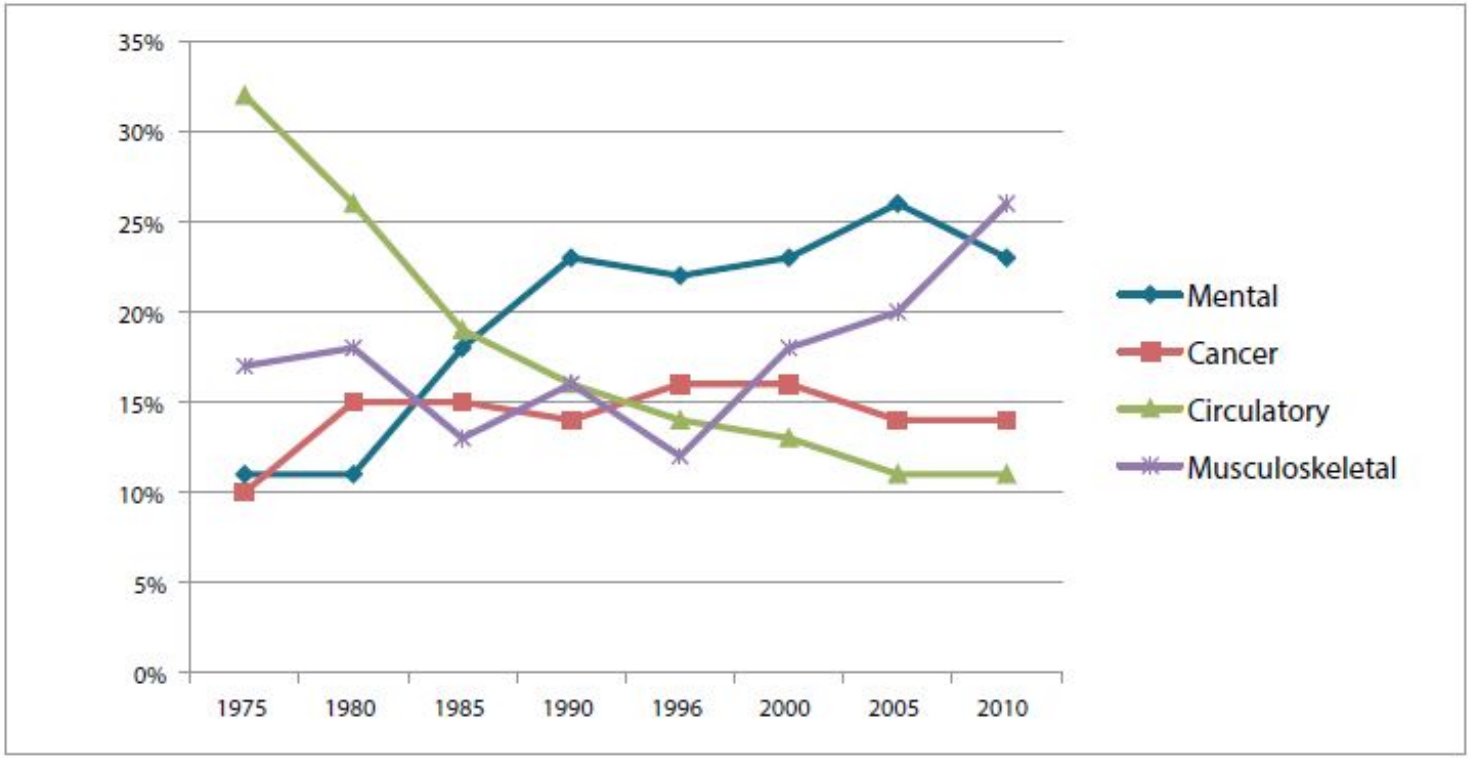


FIGURE – Cleveland (1994) recommends that curves follow a 45 ° angle. From Allen and Erhardt (2016b)

WHAT'S WRONG WITH THIS GRAPHIC ? (HARDER)

27. Initial DI Worker Awards by Major Cause of Disability—Calendar Years 1975-2010



Source: Social Security Advisory Board (2012).

FIGURE – Major Cause of Disability - 1975-2010 (J. Schwabish, 2014).

What's wrong with this graphic ? (Harder)

Questions :

- in 2010, what is the major cause of disability ?

What's wrong with this graphic ? (Harder)

Questions :

- in 2010, what is the major cause of disability ?
- in 1975, what was the major cause of disability ?

What's wrong with this graphic ? (Harder)

Questions :

- in 2010, what is the major cause of disability ?
- in 1975, what was the major cause of disability ?
- In the recent years, which causes have increased/decreased the most ?

What's wrong with this graphic ? (Harder)

Questions :

- in 2010, what is the major cause of disability ?
- in 1975, what was the major cause of disability ?
- In the recent years, which causes have increased/decreased the most ?
- . . .

What's wrong with this graphic ? (Harder)

Questions :

- in 2010, what is the major cause of disability ?
- in 1975, what was the major cause of disability ?
- In the recent years, which causes have increased/decreased the most ?
- ...
- You do not remember a damn thing of this graph !

More (smaller) graphs, are often better

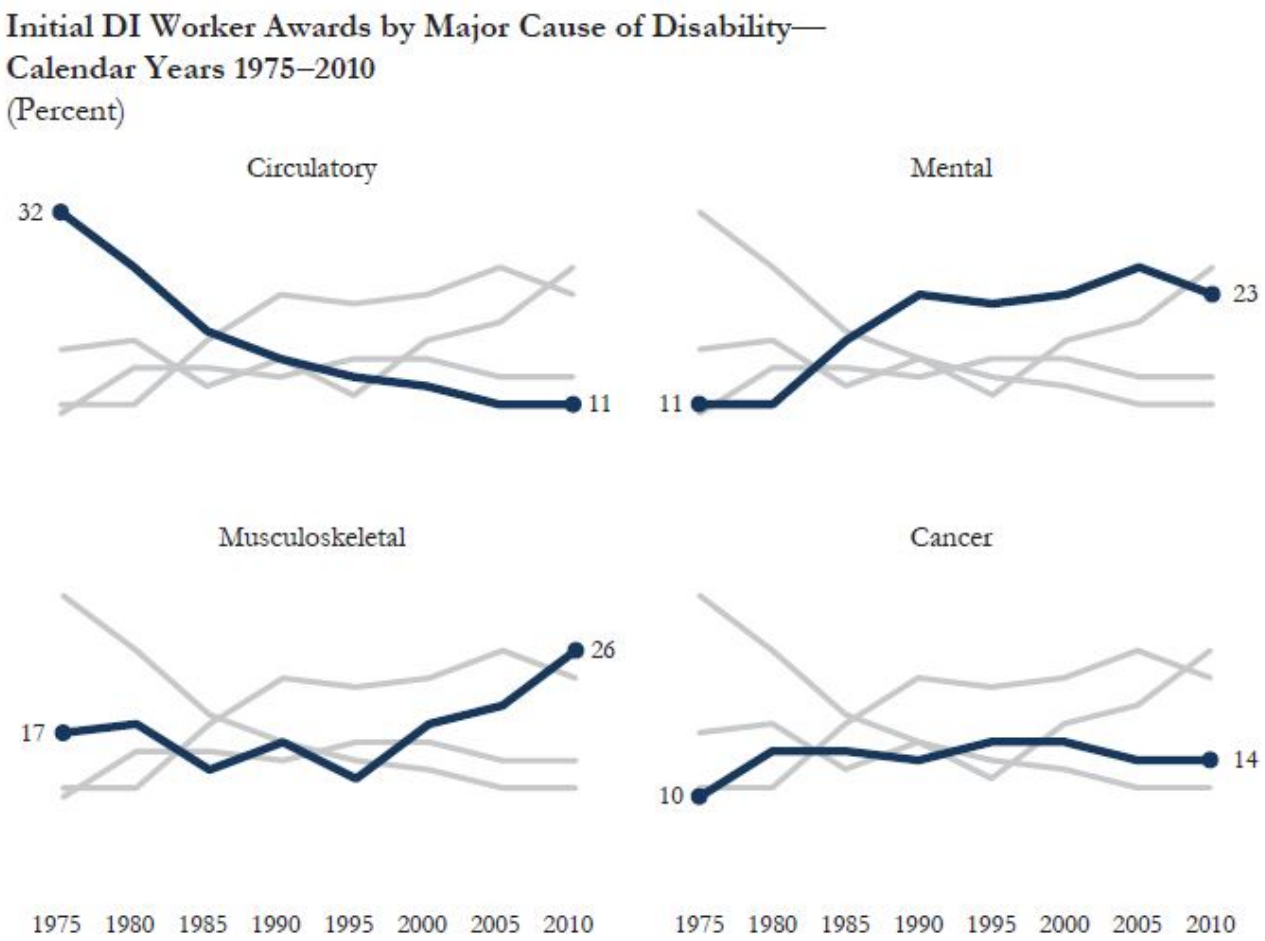
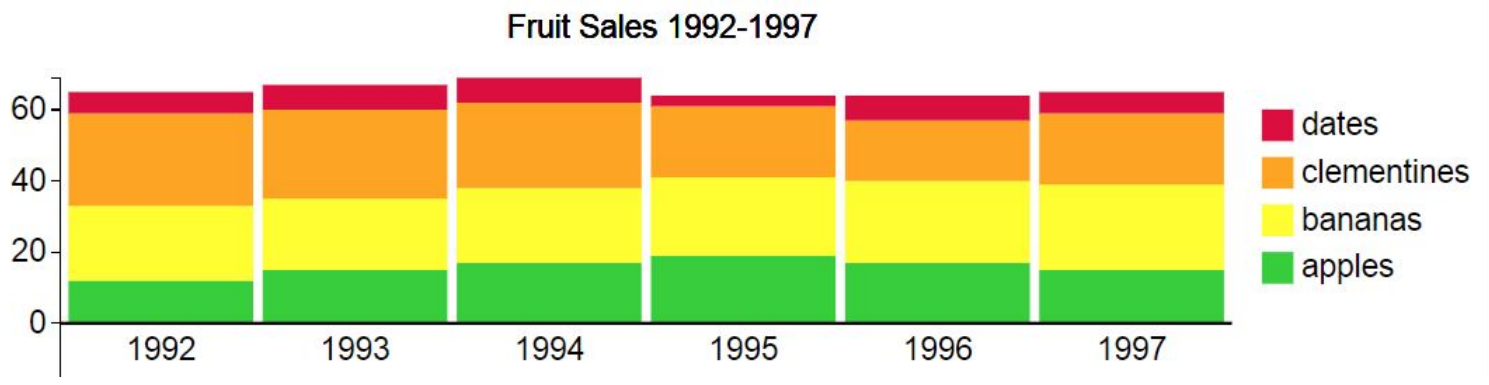


Figure – Major Cause of Disability- 1975-2010 (J. Schwabish).

Cf. "*brushing*" (ex : for parallel Coordinates plots)

YOUR TURN : WHAT'S WRONG WITH THIS GRAPHIC ?



"SMALL MULTIPLES" PROVIDE A BETTER VIEW

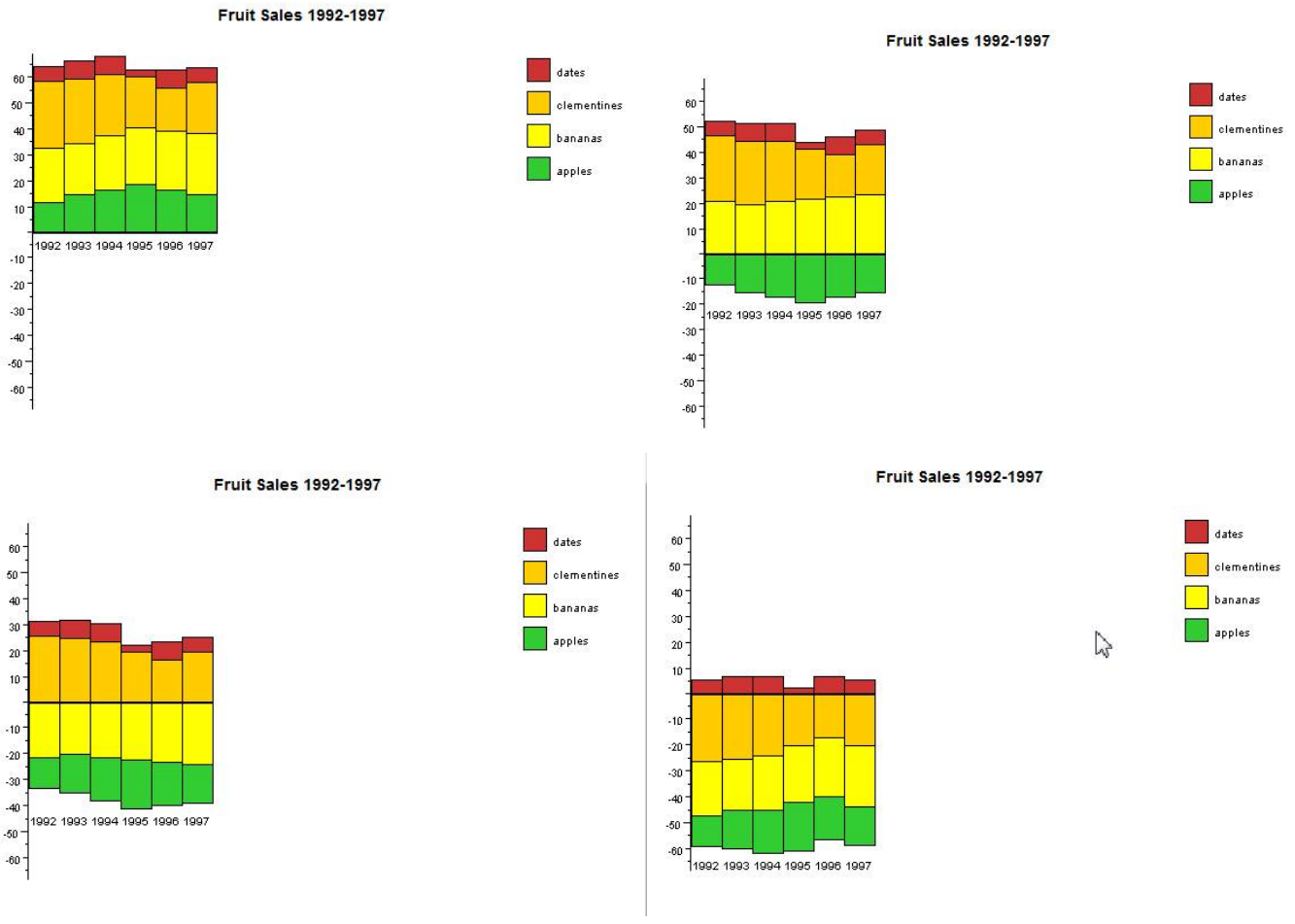


FIGURE – from Dix and Ellis (1998) example

AND HERE COMES DYNAMIC DATA VISUALISATION !

Happily Jane is using an **interactive stacked histogram**.

Try it yourself.

Click on the coloured banana entry in the key, or one of the banana parts of any of the histogram bars.

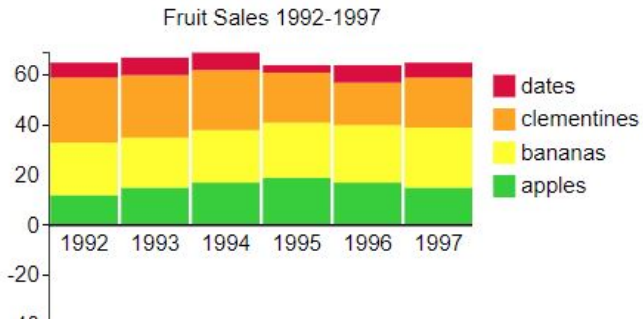
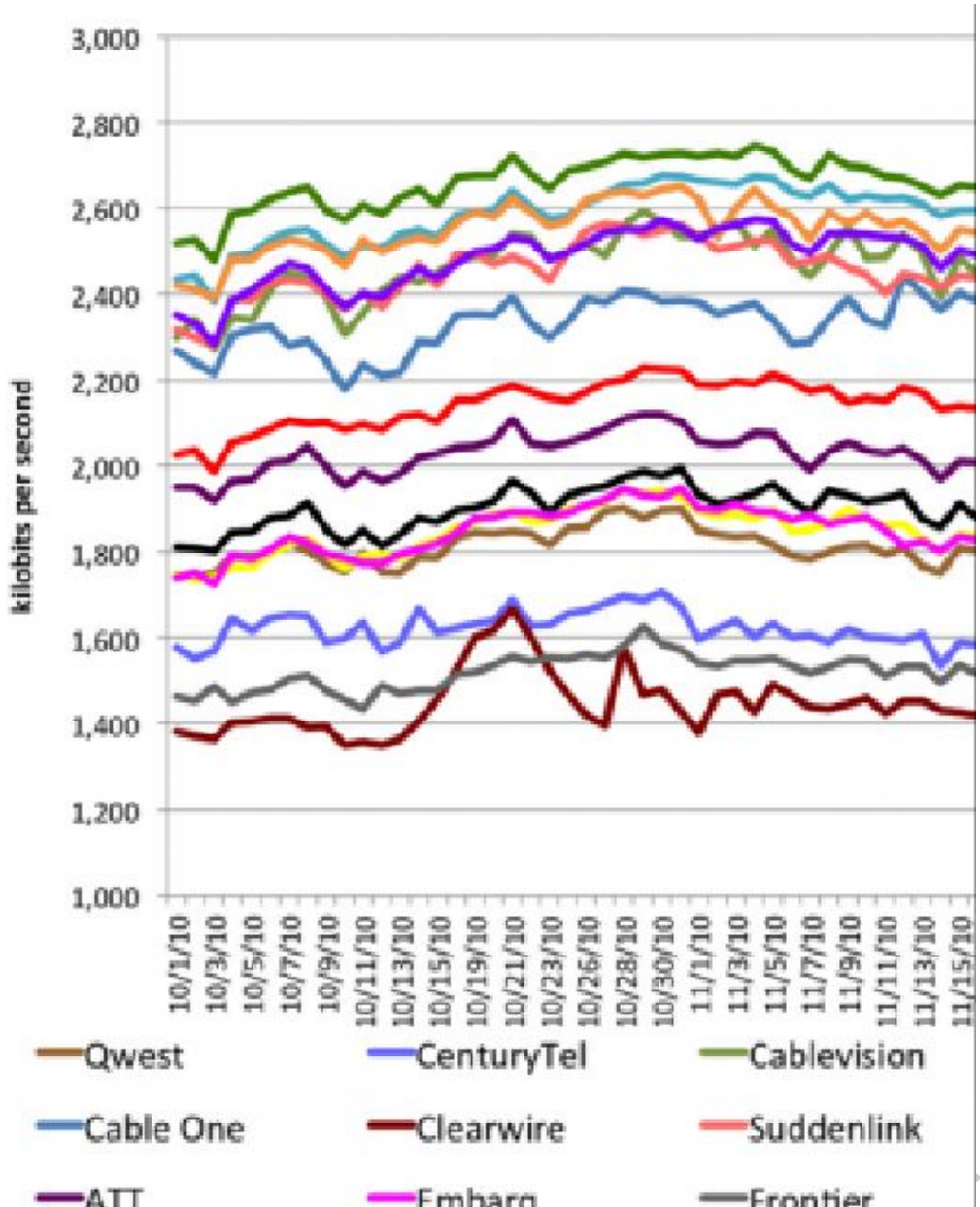


FIGURE – Dynamic fruit sales

from meandeviation.com

WHAT'S WRONG WITH THIS GRAPHIC ? (HARDER)

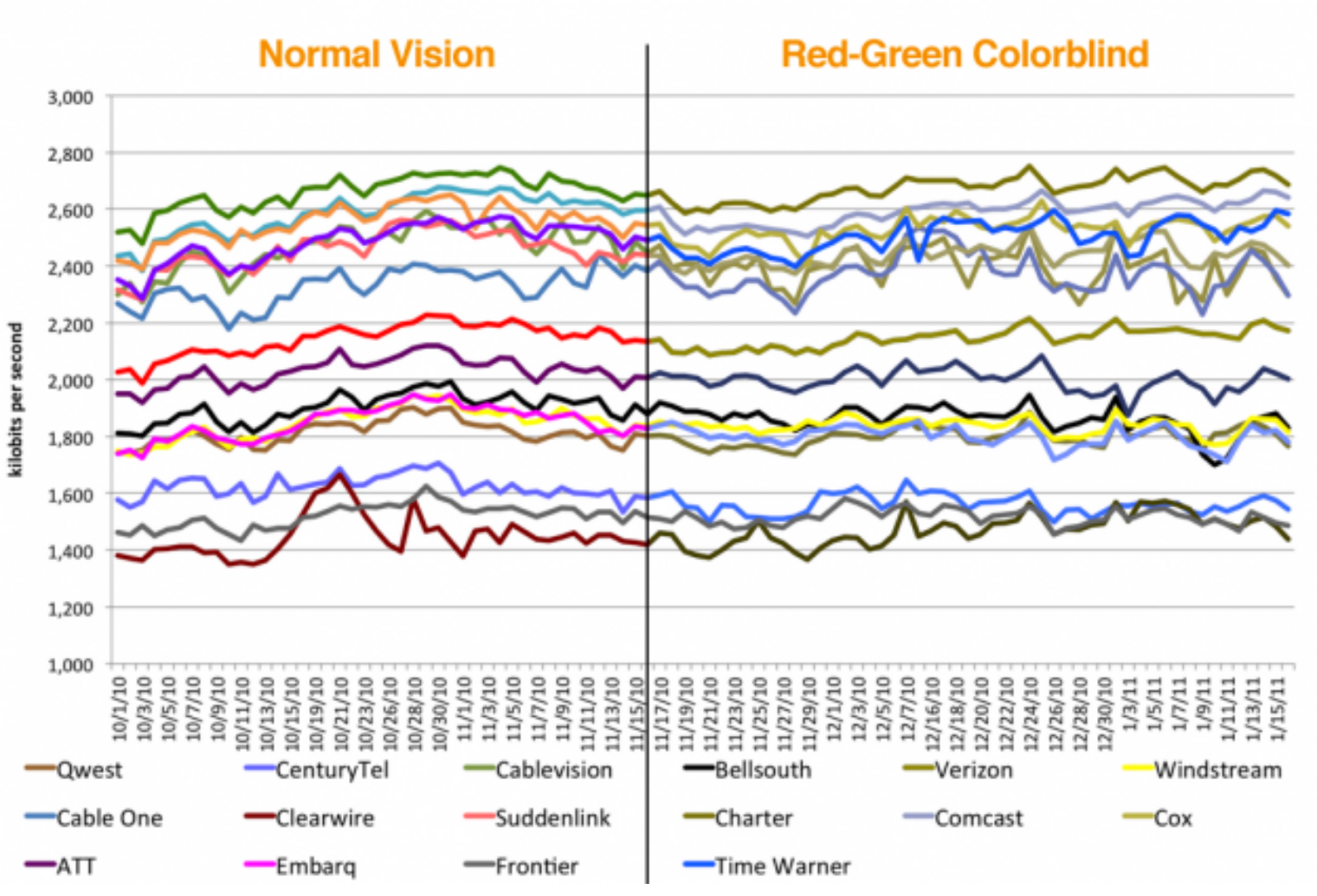


KEEP ALL YOUR AUDIENCE

Normal ↪

Color-blind ↪

WHICH MEANS THAT FOR 5 % OF MEN :



See also the ggplot option `+ scale_colour_colorblind()`

So what are the rules ?

Can you name some rules for a good (*resp*. bad) graphic ?

- Cite some rules...

So what are the rules ?

Can you name some rules for a good (*resp*. bad) graphic ?

- Cite some rules...
- Axis and scale (starting at zero !)

So what are the rules ?

Can you name some rules for a good (*resp*. bad) graphic ?

- Cite some rules...
- Axis and scale (starting at zero !)
- Context, labels

So what are the rules ?

Can you name some rules for a good (*resp*. bad) graphic ?

- Cite some rules...
- Axis and scale (starting at zero !)
- Context, labels
- No multiple scales !

So what are the rules ?

Can you name some rules for a good (*resp*. bad) graphic ?

- Cite some rules...
- Axis and scale (starting at zero !)
- Context, labels
- No multiple scales !
- Colors that are distinguishable

So what are the rules ?

Can you name some rules for a good (*resp*. bad) graphic ?

- Cite some rules...
- Axis and scale (starting at zero !)
- Context, labels
- No multiple scales !
- Colors that are distinguishable
- A good message ! Unambiguous message

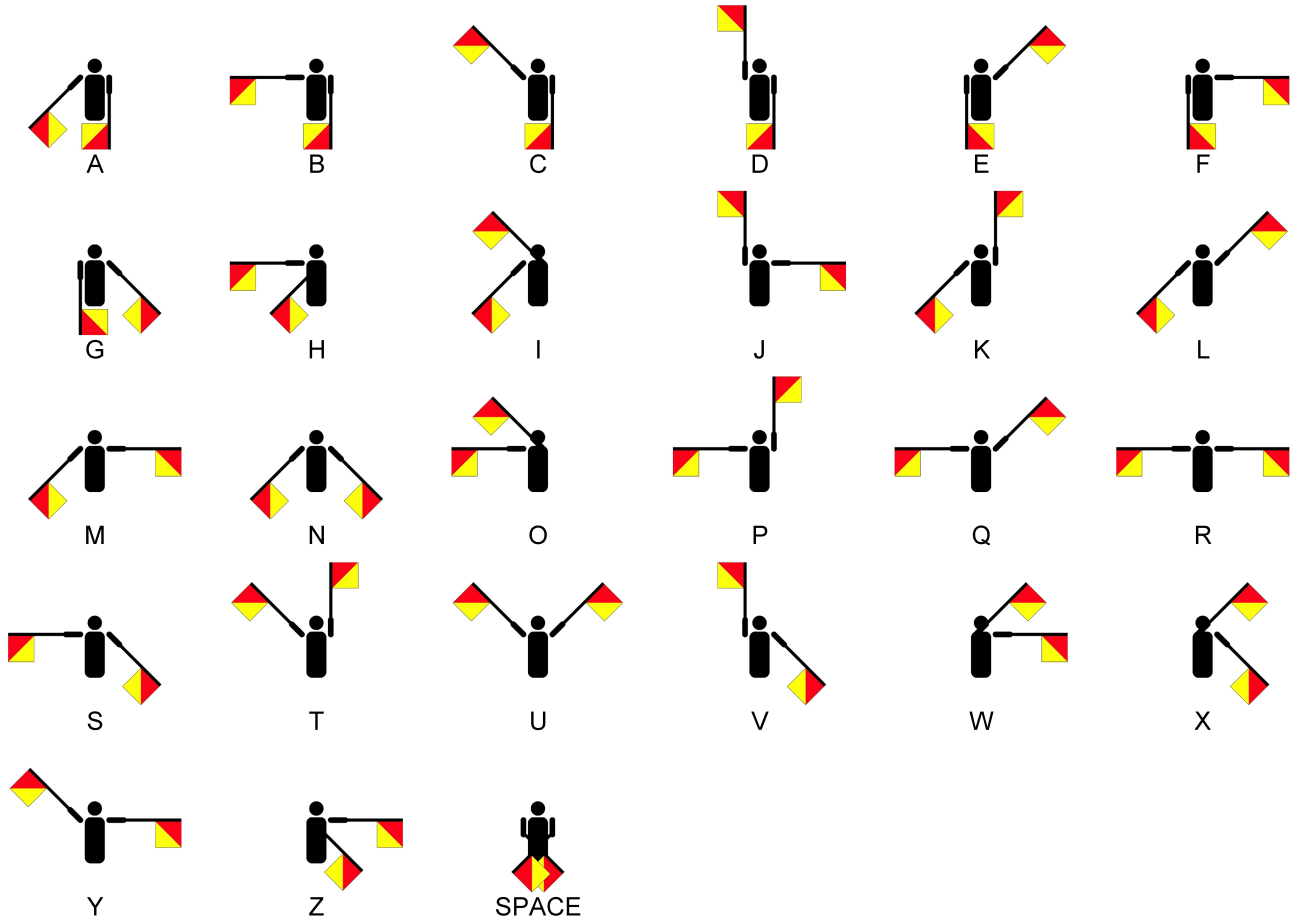
So what are the rules ?

Can you name some rules for a good (*resp*. bad) graphic ?

- Cite some rules...
- Axis and scale (starting at zero !)
- Context, labels
- No multiple scales !
- Colors that are distinguishable
- A good message ! Unambiguous message
- Should not be overloaded

[- WHY CODING ? -]

If decoding is hard, why coding ?



GRAPHICS *reveal* DATA : ANSCOMBE (1973) QUARTET

We use here 4 couples of random variables : (X_1, Y_1), (X_2, Y_2) (X_3, Y_3) and (X_4, Y_4). All four data sets have the same descriptive statistics.

Xs	**Mean**	**Std. Dev.**	**Ys**	**Mean**	**Std. Dev.**	$corr(X_i, Y_i)$	**N**
X_1	9	3.32	Y_1	7.5	2.03	0.8164	11
X_2	9	3.32	Y_2	7.5	2.03	0.8162	11
X_3	9	3.32	Y_3	7.5	2.03	0.8163	11
X_4	9	3.32	Y_4	7.5	2.03	0.8165	11

ANSCOMBE (1973) QUARTET

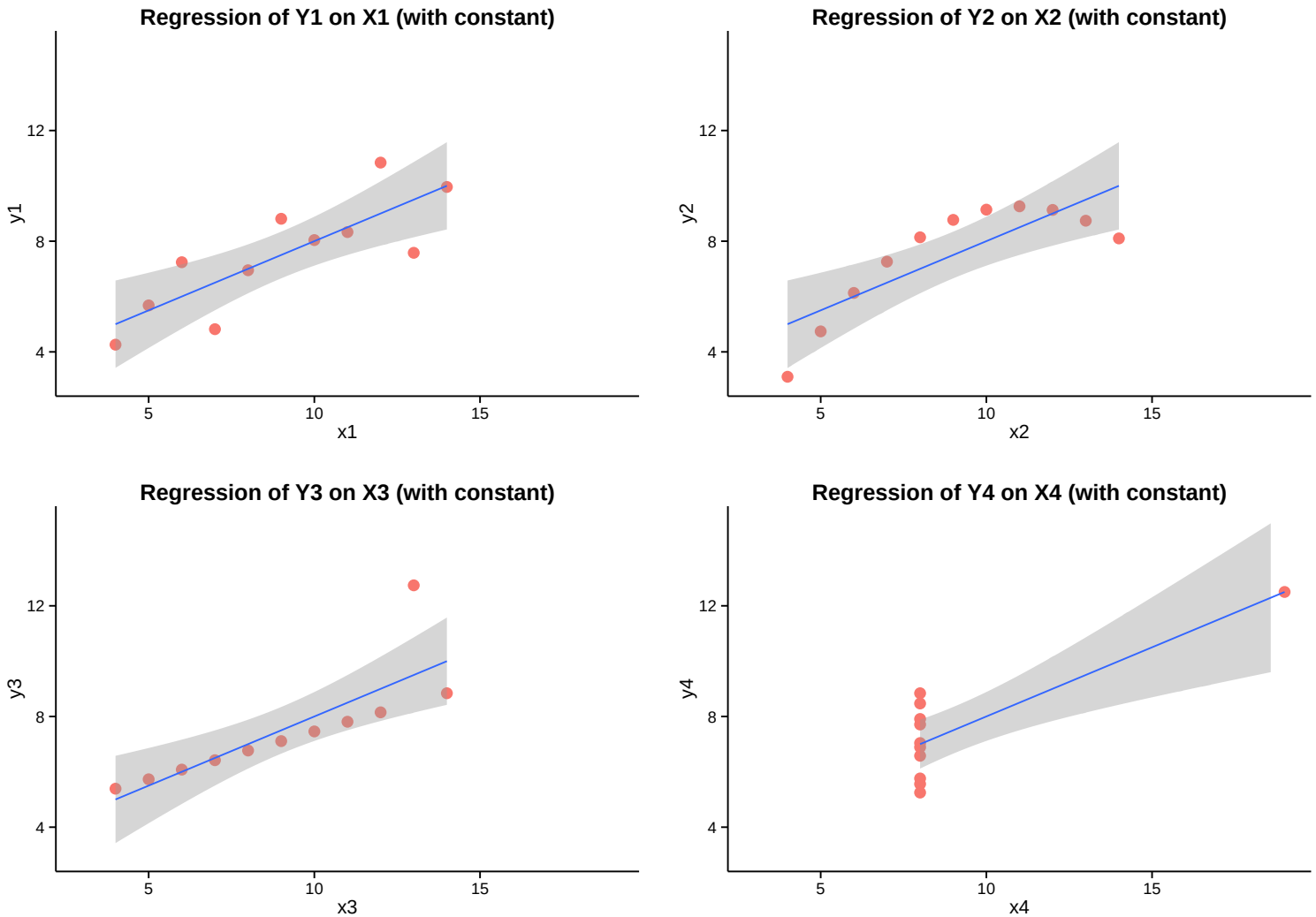
All four data sets are described by the same linear model ($Y_i = \alpha + \beta X_i + \epsilon_i$), revealing apparently the same relationships :

	Dependent variable :			
	Y_1	Y_2	Y_3	Y_4
Regressed on :				
X_i, i=1,...,4	0.500 ***	0.500***	0.500***	0.500***
Constant	3.000**	3.001**	3.002**	3.002**
R^2	0.667	0.666	0.666	0.667
Resid Std. Error	1.237	1.237	1.236	1.236
F Statistic	17.990***	17.966***	17.972***	18.003***

Note : Data from Anscombe (1973). * p <0.1 ; ** p < 0.05 ; *** p < 0.01

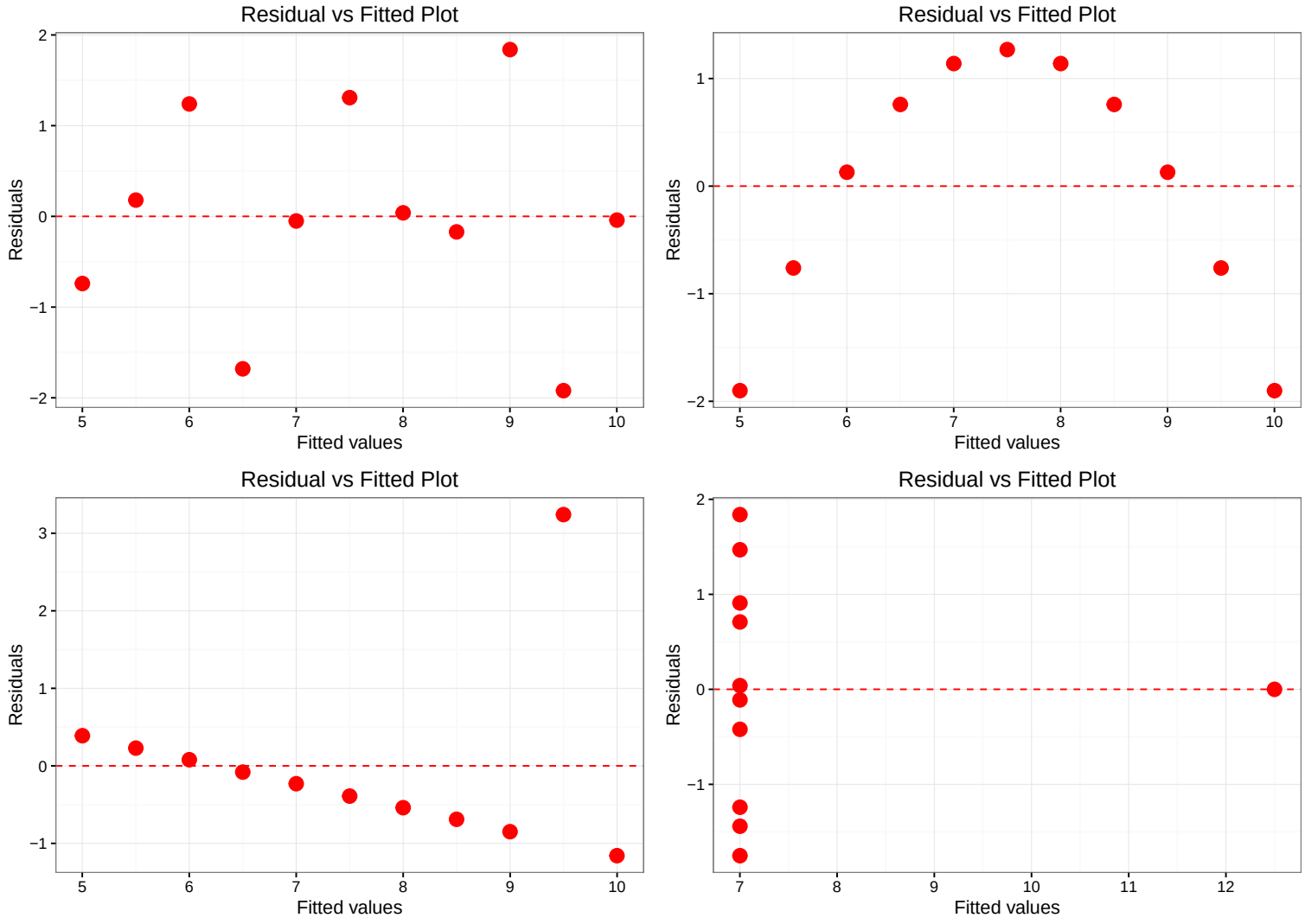
Anscombe (1973) quartet

A simple scatter plot (regression overlaid) shows something very different.

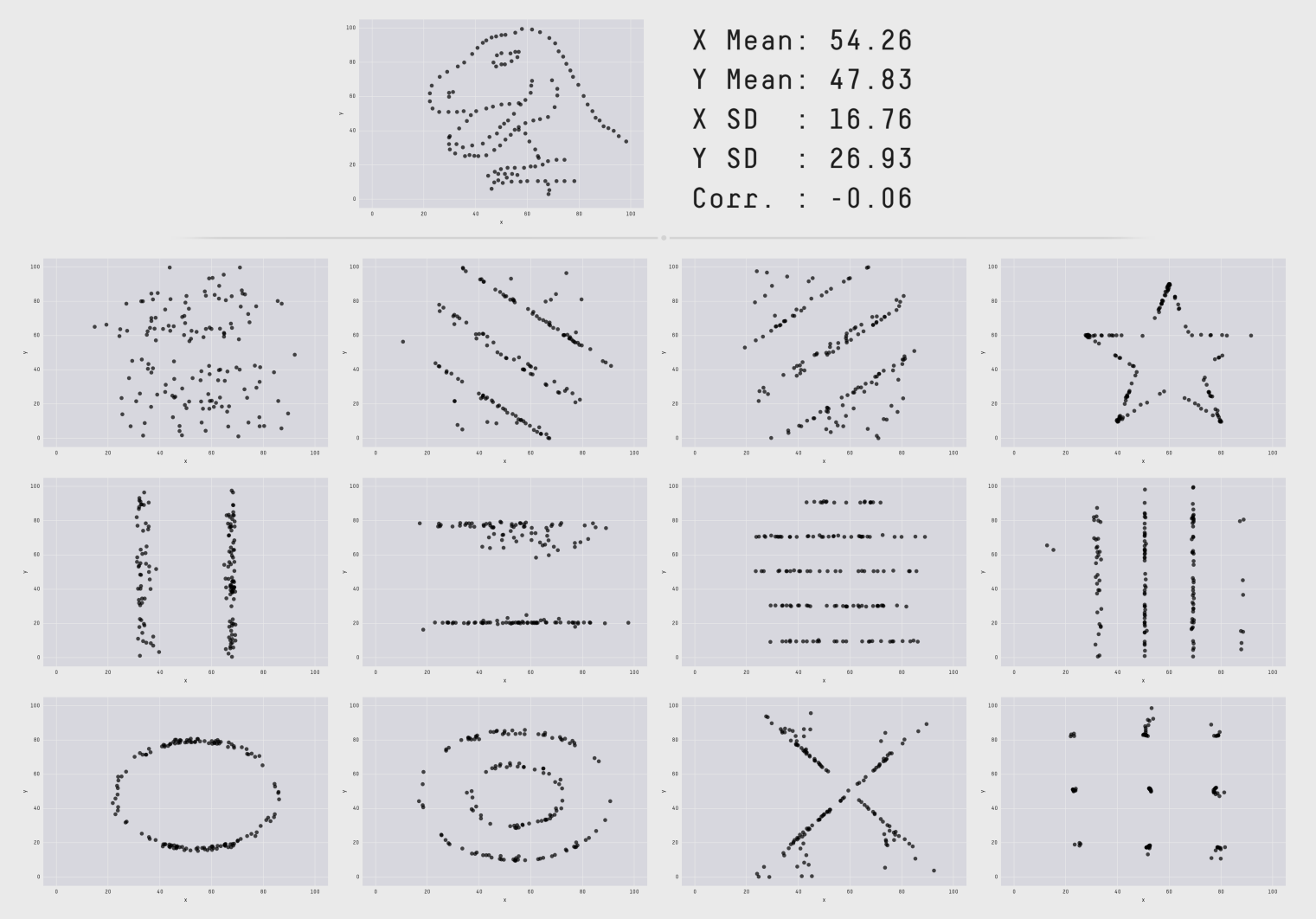


ANSCOMBE (1973) QUARTET

NP : Plots of the residuals shows also same differences



The Datasaurus...



Matejka and Fitzmaurice (2017), see also The Datasaurus website

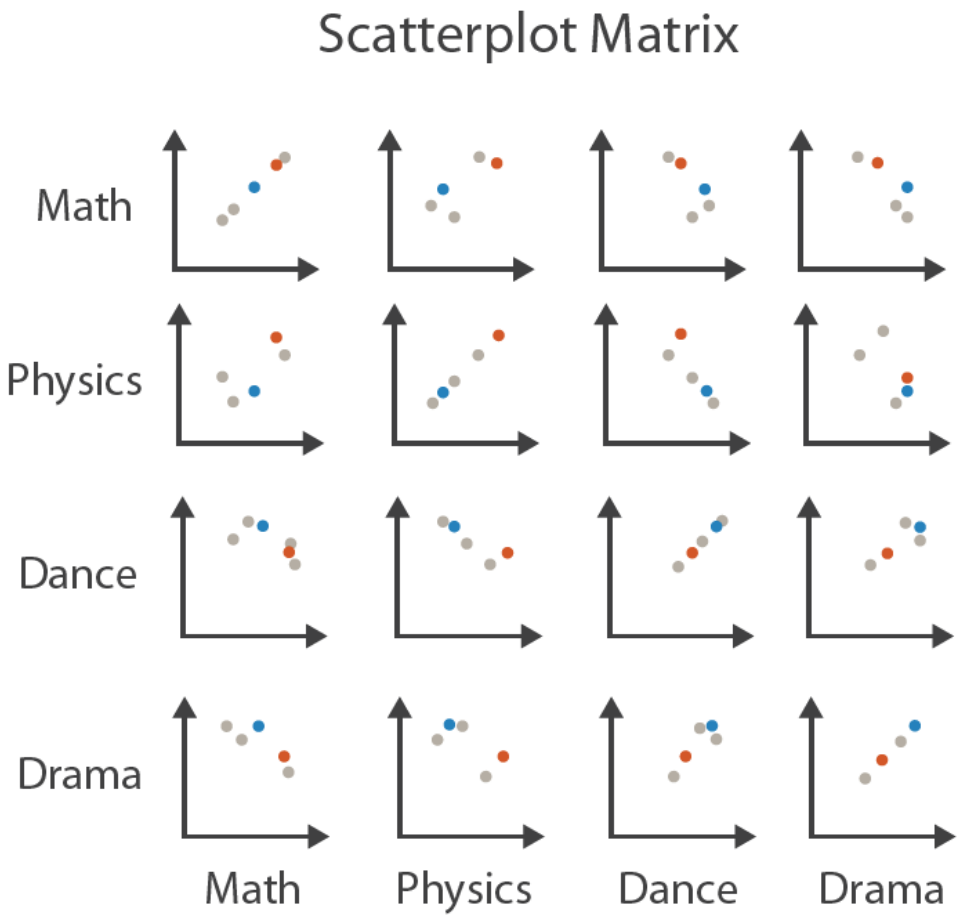
TABLES VS GRAPHICS

Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

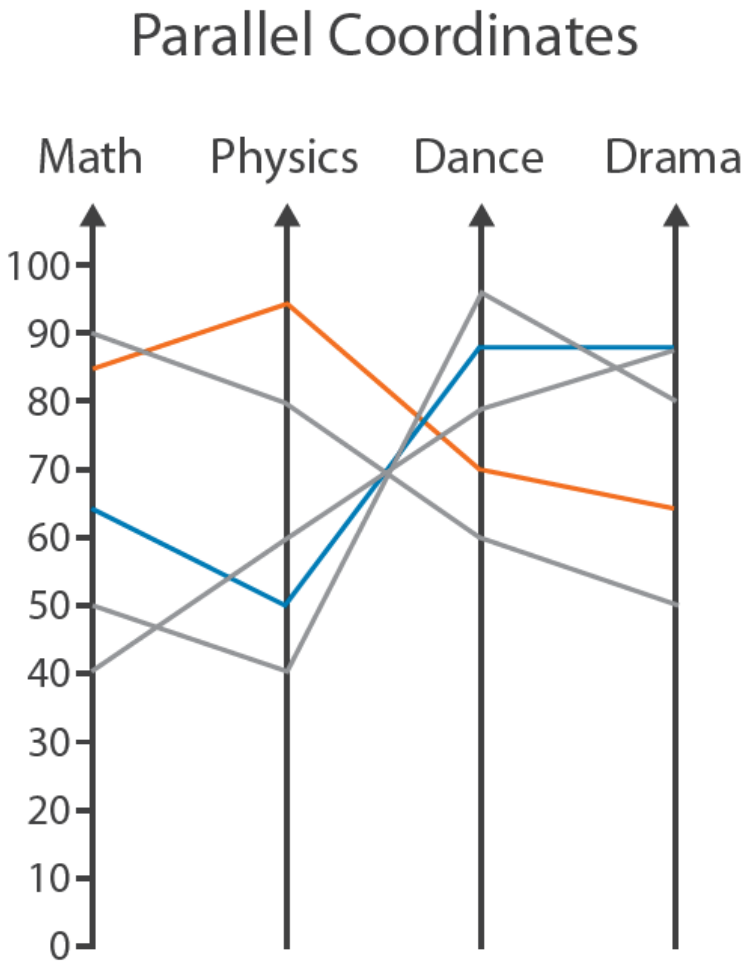
From Munzner (2014)

TABLES VS GRAPHICS



From Munzner (2014)

TABLES VS GRAPHICS

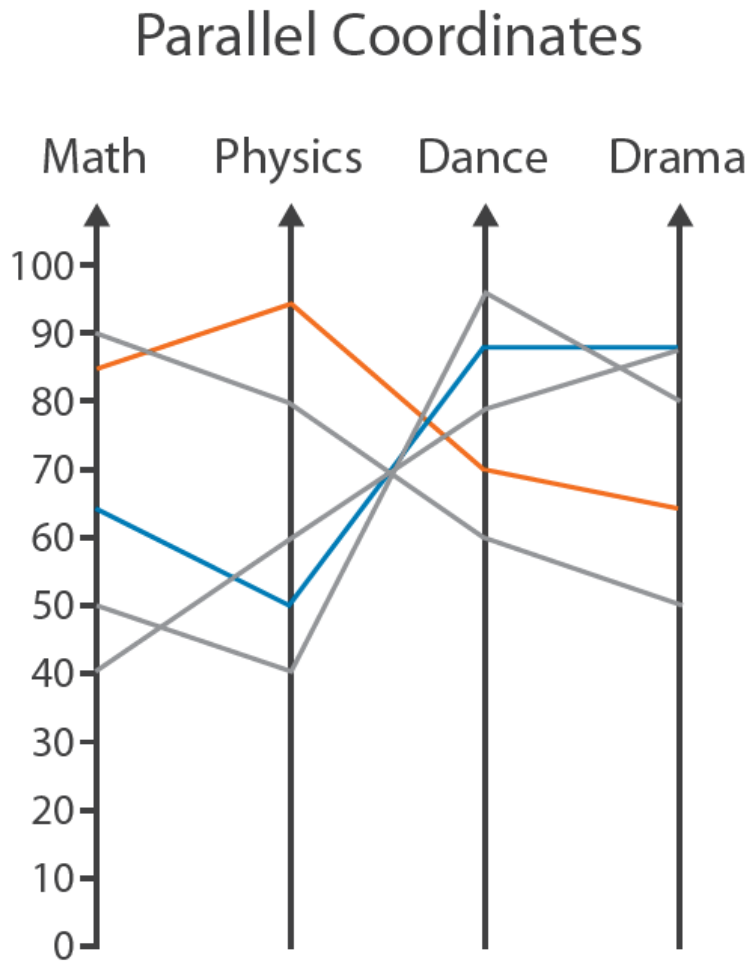


From Munzner (2014)

TABLES VS GRAPHICS

Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90



From Munzner (2014)

TABLES VS GRAPHICS : **CASE STUDY**

Data with many 0/1 variables :
Facilities indicators for 16 towns (A-P)

Cities																
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**#**	**Facilities**
0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	**1**	**School**
0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	0	**2**	**Agricultural Coop**
0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	**3**	**Station**
1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	**4**	**School (one class)**
0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	0	**5**	**Vet**
1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	**6**	**No doctor**
0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	**7**	**No running water**
0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	**8**	**Police station**
0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	0	**9**	**Land consolidation**

from Bertin (1981)

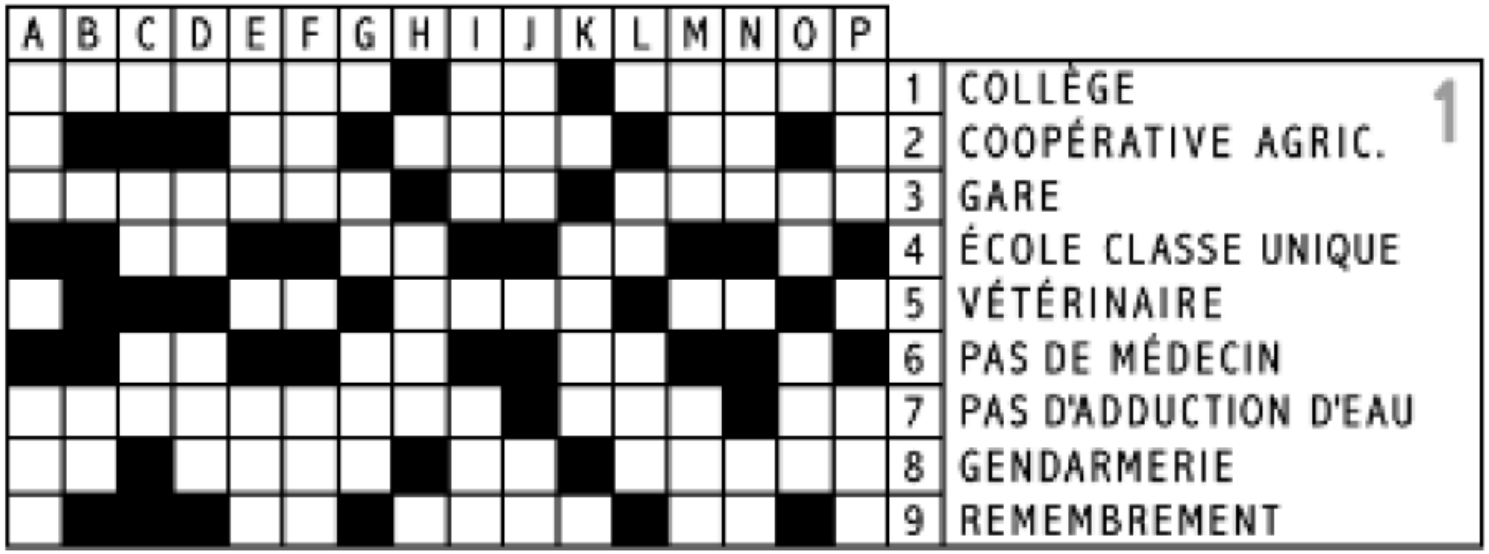
TABLES VS GRAPHICS :

"Ordering is visualising"

Using Bertin (1981) principles, The AVIZ lab has created the *Bertifier*. See also the R package Hahsler et al. (2008).

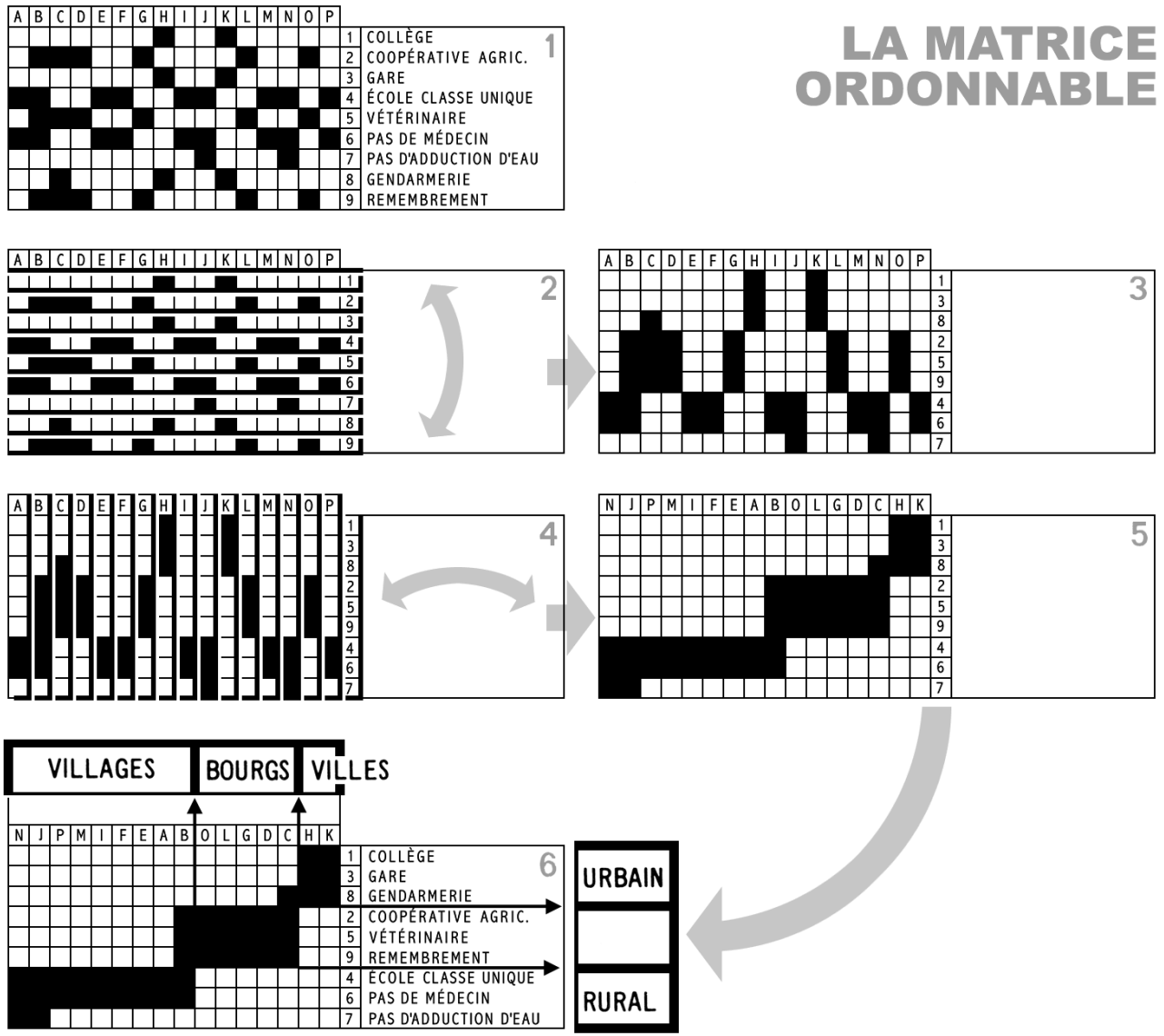
TABLES VS GRAPHICS :

Data with many 0/1 variables :
Facilities indicators for 16 towns (A-P)

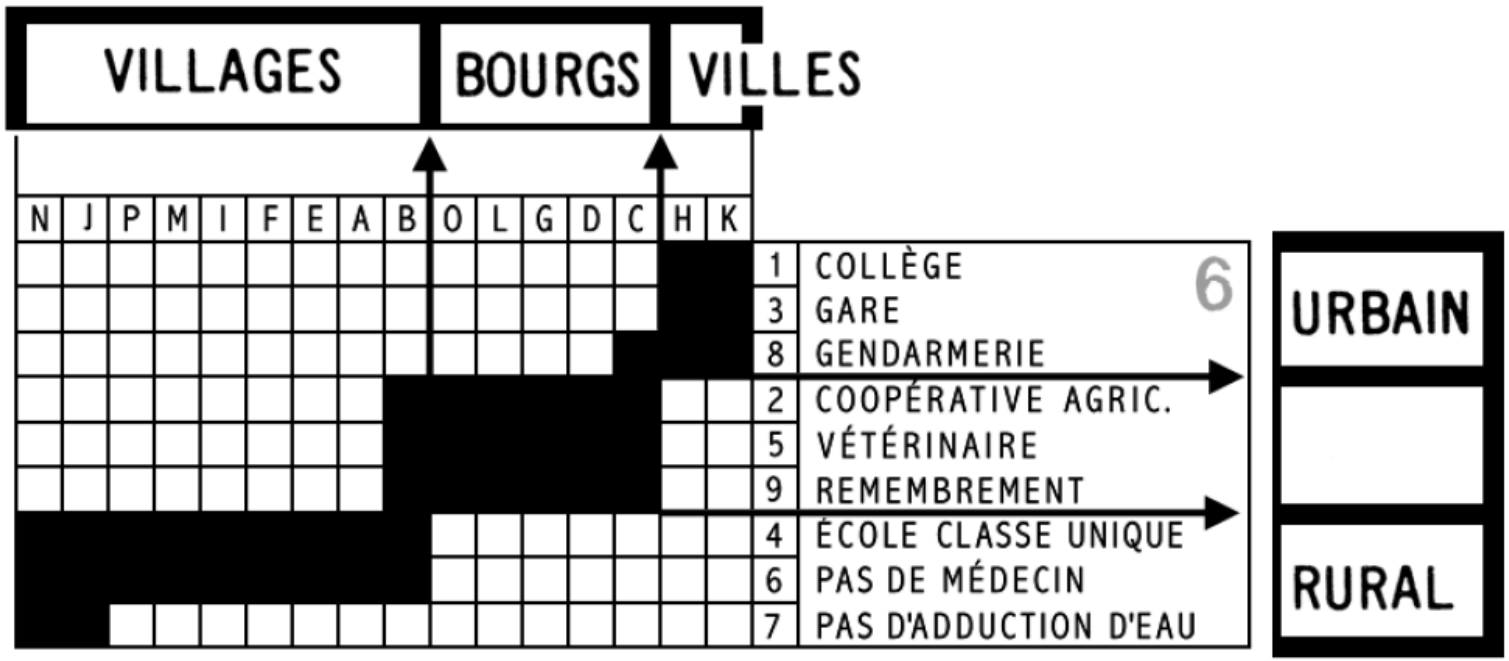


Bertin (1981)

TABLES VS GRAPHICS



TABLES VS GRAPHICS



Bertin (1981)

TABLES VS GRAPHICS

ordering is the key

N		J	N	A	E	F	I	M	P	B	D	G	L	O	C	K	H
1	HIGH SCHOOL															■	■
3	RAILWAY STATION															■	■
8	POLICE STATION														■	■	■
2	AGRICULTURAL COOPERATIVE									■	■	■	■	■	■		
5	VETERINARY									■	■	■	■	■	■		
9	LAND REALLOCATION									■	■	■	■	■	■		
4	ONE ROOM SCHOOL	■	■	■	■	■	■	■	■	■							
6	NO DOCTOR	■	■	■	■	■	■	■	■	■							
7	NO WATER SUPPLY	■															

Using Bertin (1981) principles, The AVIZ lab has created the *Bertifier*. See also the R package Hahsler et al. (2008).

GOOD GRAPHICS ?

It the excellent Handbook of data visualisation Chen et al. (2007), we find some good questions :

- What to Whom, How and Why ?
  *A graphic may be linked to three pieces of text : its **caption**, a **headline** and **an article** it accompanies. Ideally, all three should be consistent and complement each other.*

GOOD GRAPHICS ?

It the excellent Handbook of data visualisation Chen et al. (2007), we find some good questions :

- What to Whom, How and Why ?
  *A graphic may be linked to three pieces of text : its **caption**, a **headline** and **an article** it accompanies. Ideally, all three should be consistent and complement each other.*
- Present or explore data ?
  Different purpose, different requirements !

GOOD GRAPHICS ?

It the excellent Handbook of data visualisation Chen et al. (2007), we find some good questions :

- What to Whom, How and Why ?
  *A graphic may be linked to three pieces of text : its **caption**, a **headline** and **an article** it accompanies. Ideally, all three should be consistent and complement each other.*
- Present or explore data ?
  Different purpose, different requirements !
- Choice of Graphical form ?
  Choice depends on the type of data to be displayed (e.g. univariate continuous data, bivariate categorical data, etc..) and on what is to be shown.

GOOD GRAPHICS ?

It the excellent Handbook of data visualisation Chen et al. (2007), we find some good questions :

- What to Whom, How and Why ?
  *A graphic may be linked to three pieces of text : its **caption**, a **headline** and **an article** it accompanies. Ideally, all three should be consistent and complement each other.*
- Present or explore data ?
  Different purpose, different requirements !
- Choice of Graphical form ?
  Choice depends on the type of data to be displayed (e.g. univariate continuous data, bivariate categorical data, etc..) and on what is to be shown.
- Unique solution ?
  *There is **not** always a unique optimal choice and alternatives can be equally good or good in different ways, emphasizing different aspects of the same data.*

"Good" or "bad" Graphics ?

"There are no "good" nor "bad" graphics (...), there are graphics answering legitimate questions and graphics that do not answer question at all "

Bertin (1981)

It is easy to criticize ... but are there some rules ?

WHAT'S WRONG WITH THIS GRAPHIC ?

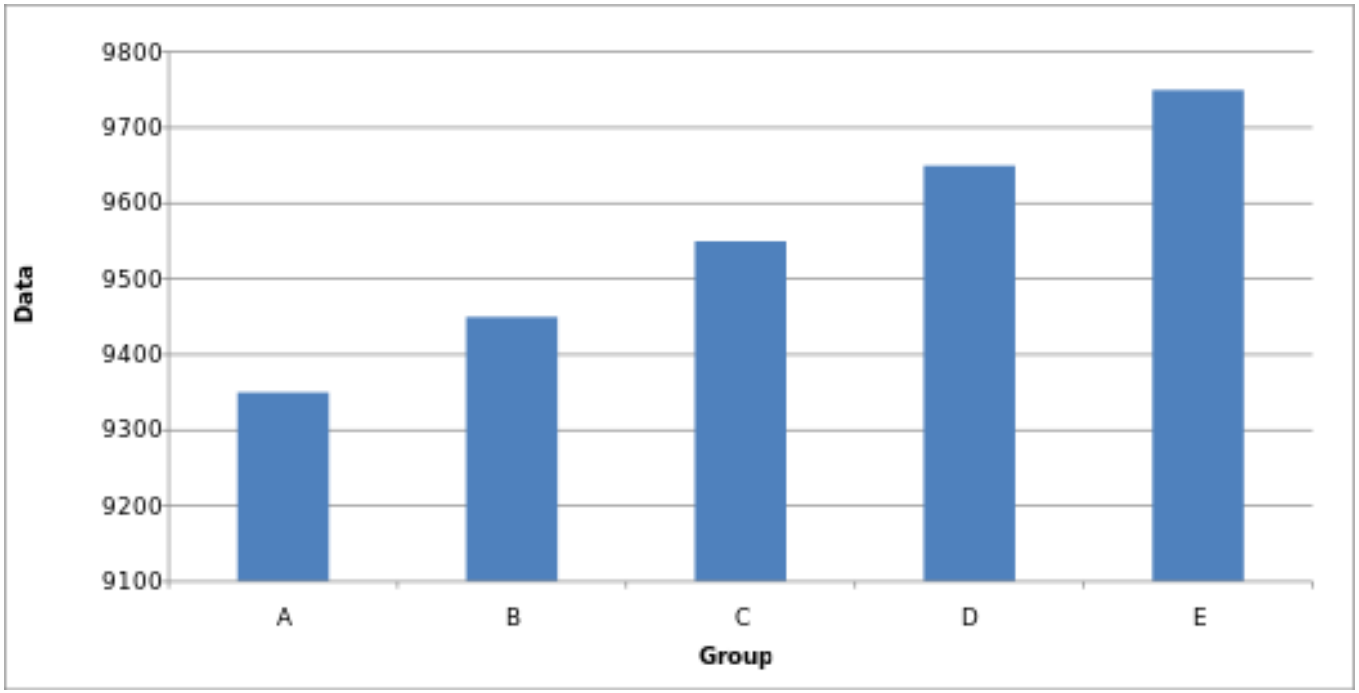


FIGURE – (source Wikipedia)

WHAT'S WRONG WITH THIS GRAPHIC ?

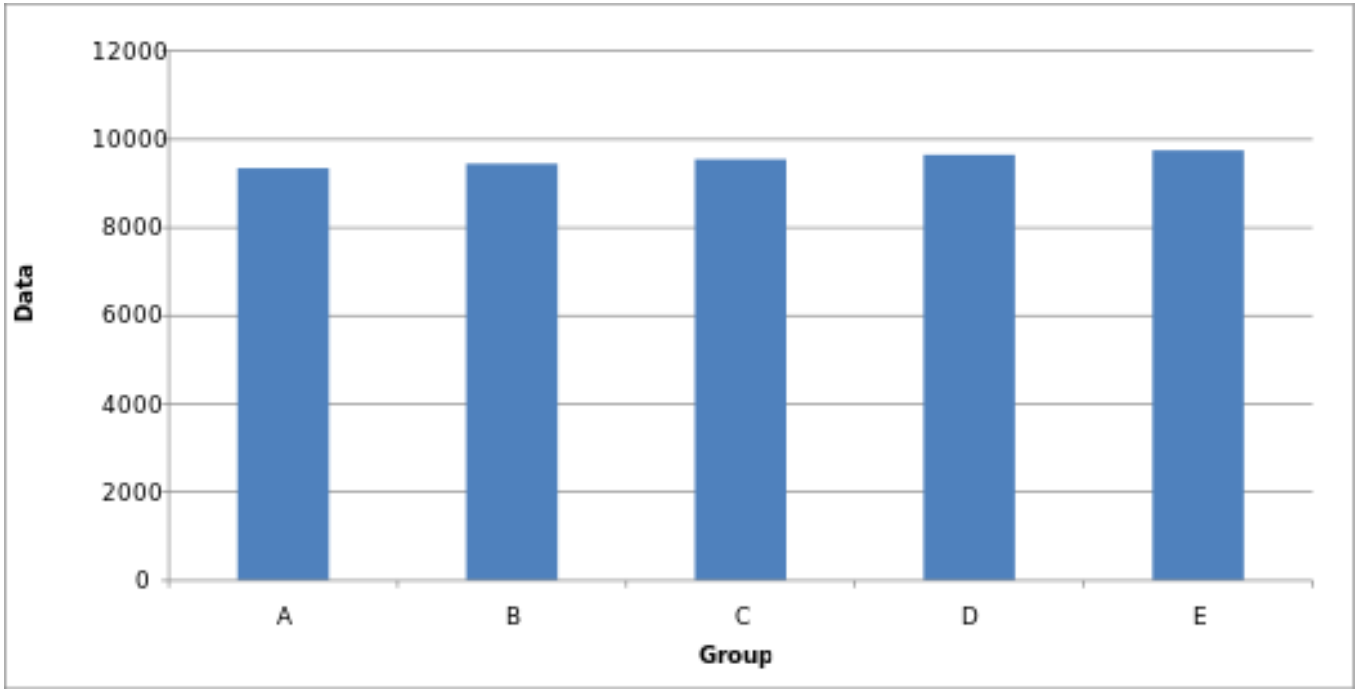


FIGURE – (source Wikipedia)

WHAT'S WRONG WITH THIS GRAPHIC ? (HARDER)

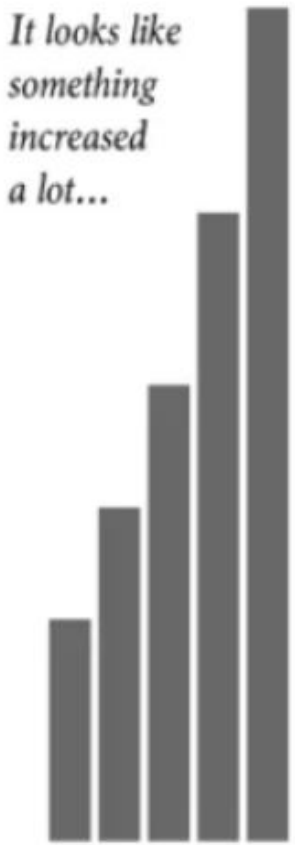


FIGURE – Are you looking at the right thing ?

WHAT'S WRONG WITH THIS GRAPHIC ? (HARDER)

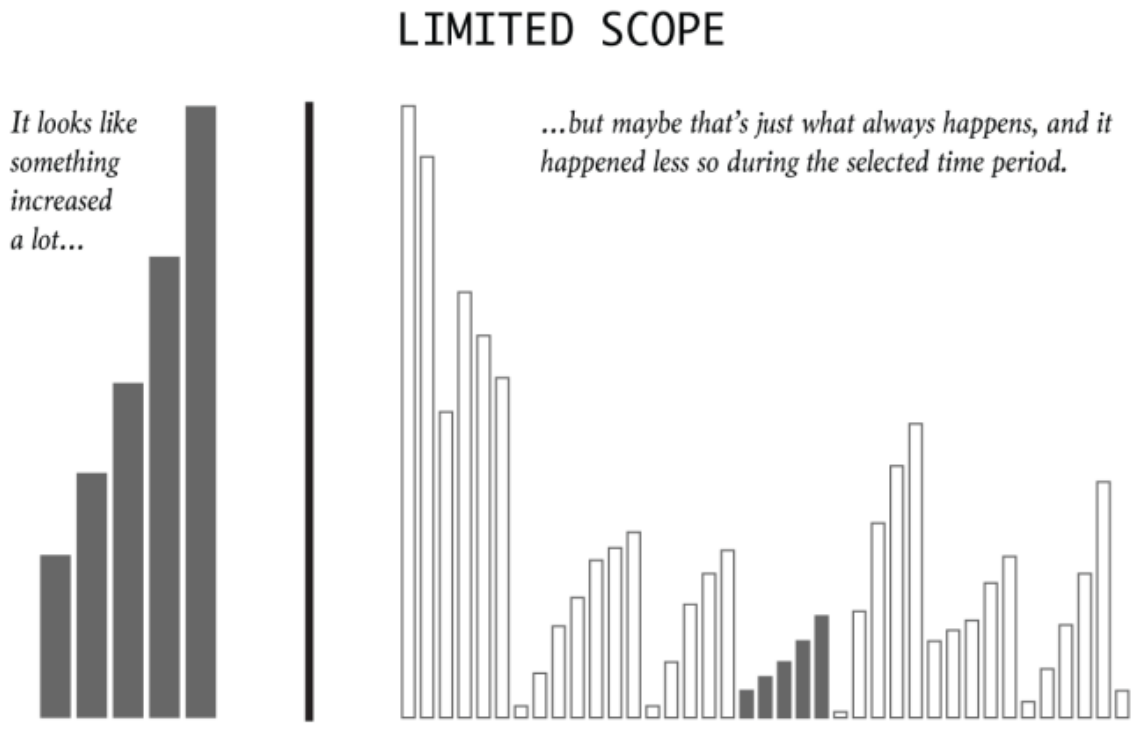


FIGURE – Are you looking at the right thing ?

from Flowing data

Edward R. Tufte's Rules

In his seminal book, Tufte (2001) propose some principles for displaying quantitative information.

Data : *Above all,* ***show the data***

Edward R. Tufte's Rules

In his seminal book, Tufte (2001) propose some principles for displaying quantitative information.

Data : *Above all,* ***show the data***

Question : *Induce the viewer to* ***think about the substance*** *rather than about methodology, graphic design. Encourage the eye to compare different piece of data.*

Edward R. Tufte's Rules

In his seminal book, Tufte (2001) propose some principles for displaying quantitative information.

Data : *Above all,* ***show the data***

Question : *Induce the viewer to* ***think about the substance*** *rather than about methodology, graphic design. Encourage the eye to compare different piece of data.*

Data-ink ratio : *Maximize the* ***ink-data ratio****. Erase all non data ink, Erase redundant information*

Edward R. Tufte's Rules

In his seminal book, Tufte (2001) propose some principles for displaying quantitative information.

Data : *Above all,* ***show the data***

Question : *Induce the viewer to* ***think about the substance*** *rather than about methodology, graphic design. Encourage the eye to compare different piece of data.*

Data-ink ratio : *Maximize the* ***ink-data ratio****. Erase all non data ink, Erase redundant information*

Integrity : ***Avoid distorting*** *what the data have to say*

EDWARD R. TUFTE'S RULES

In his seminal book, Tufte (2001) propose some principles for displaying quantitative information.

Data : *Above all,* ***show the data***

Question : *Induce the viewer to* ***think about the substance*** *rather than about methodology, graphic design. Encourage the eye to compare different piece of data.*

Data-ink ratio : *Maximize the* ***ink-data ratio****. Erase all non data ink, Erase redundant information*

Integrity : ***Avoid distorting*** *what the data have to say*

General to specific : *Reveal the data at* ***different levels of detail*** *(from broad picture to fine structure)*

EDWARD R. TUFTE'S RULES

In his seminal book, Tufte (2001) propose some principles for displaying quantitative information.

Data : *Above all,* ***show the data***

Question : *Induce the viewer to* ***think about the substance*** *rather than about methodology, graphic design. Encourage the eye to compare different piece of data.*

Data-ink ratio : *Maximize the* ***ink-data ratio****. Erase all non data ink, Erase redundant information*

Integrity : ***Avoid distorting*** *what the data have to say*

General to specific : *Reveal the data at* ***different levels of detail*** *(from broad picture to fine structure)*

Context : *Graphical display should be closely integrated with the* ***statistical and verbal descriptions*** *of the data set.*

PRACTICAL EXAMPLE : DATA-INK RATIO

Let's start with a classical graph (R default - `Boxplot`)

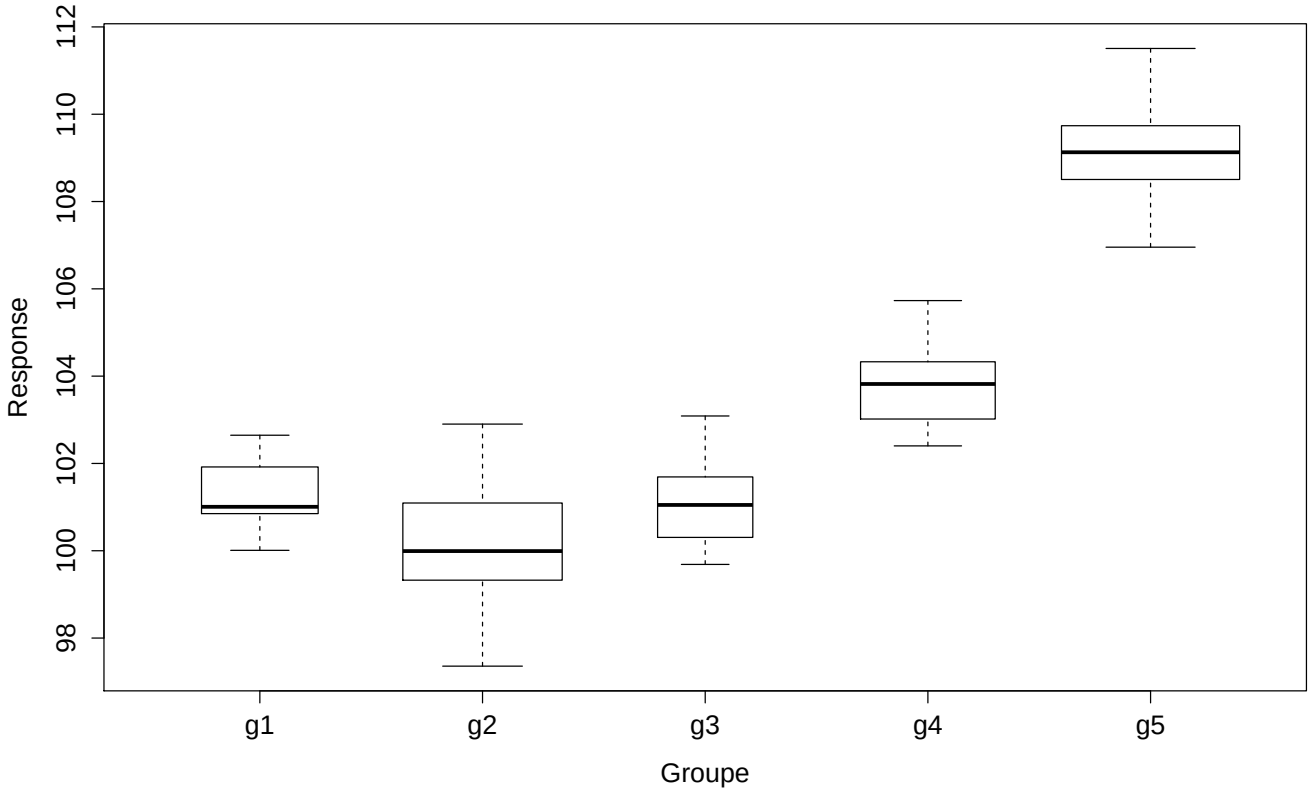


FIGURE – Distribution of a continuous variable on 4 groups

ERASE ALL NON DATA INK

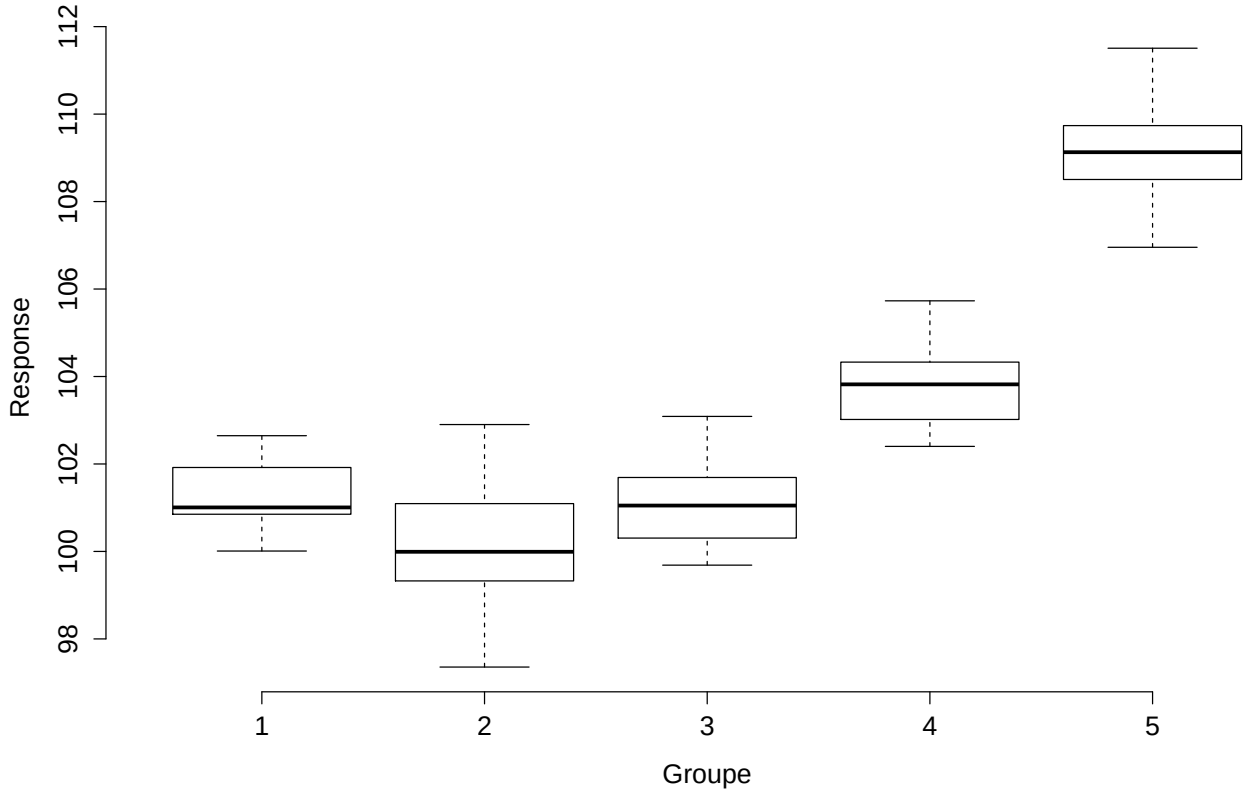


FIGURE – Distribution of a continuous variable on 4 groups

ERASE ALL REDUNDANT !

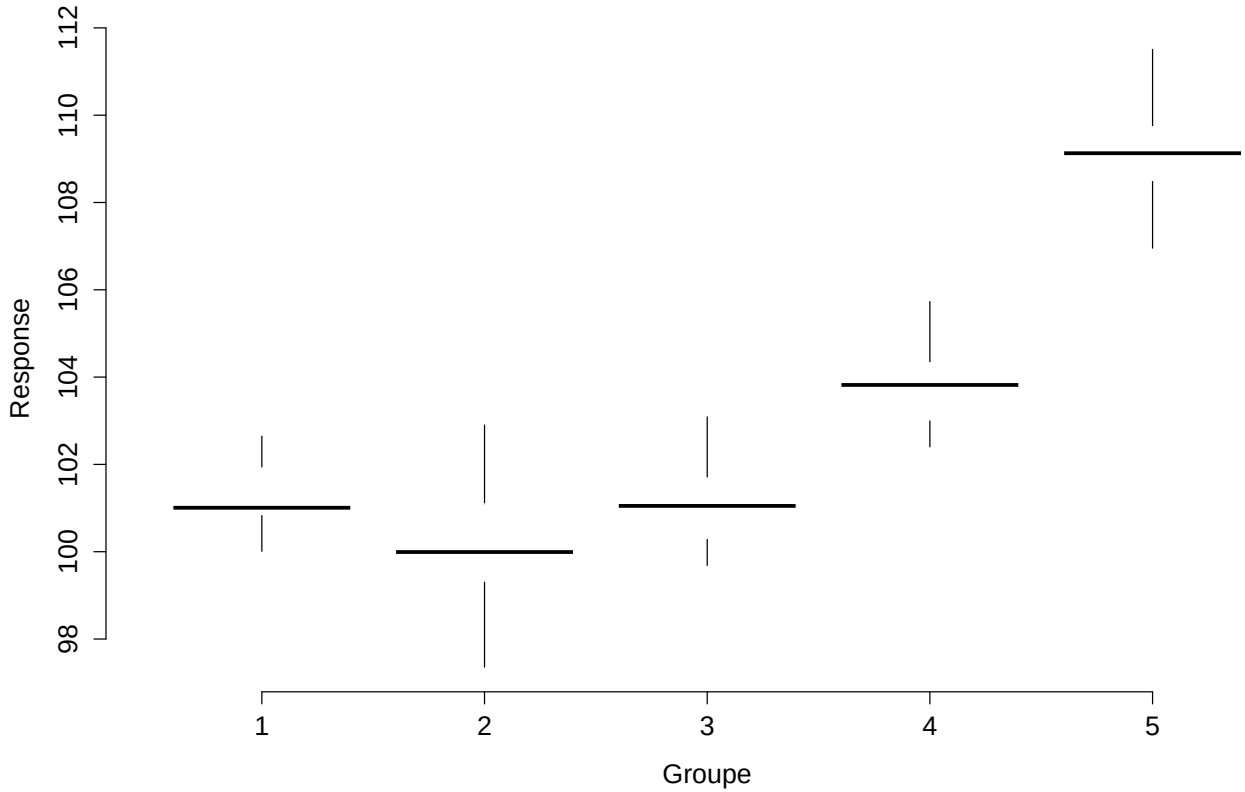


FIGURE – Distribution of a continuous variable on 4 groups

GOING FURTHER...

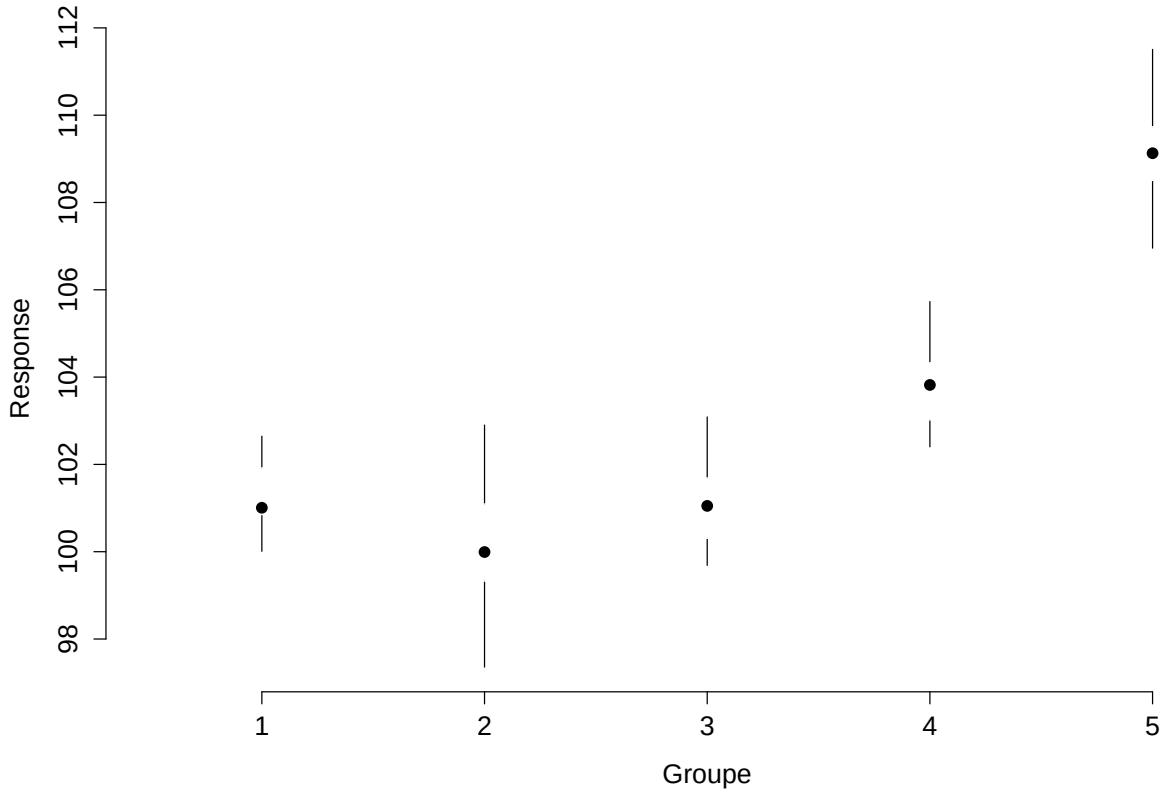


FIGURE – Distribution of a continuous variable on 4 groups

AND SHOW THE DATA...

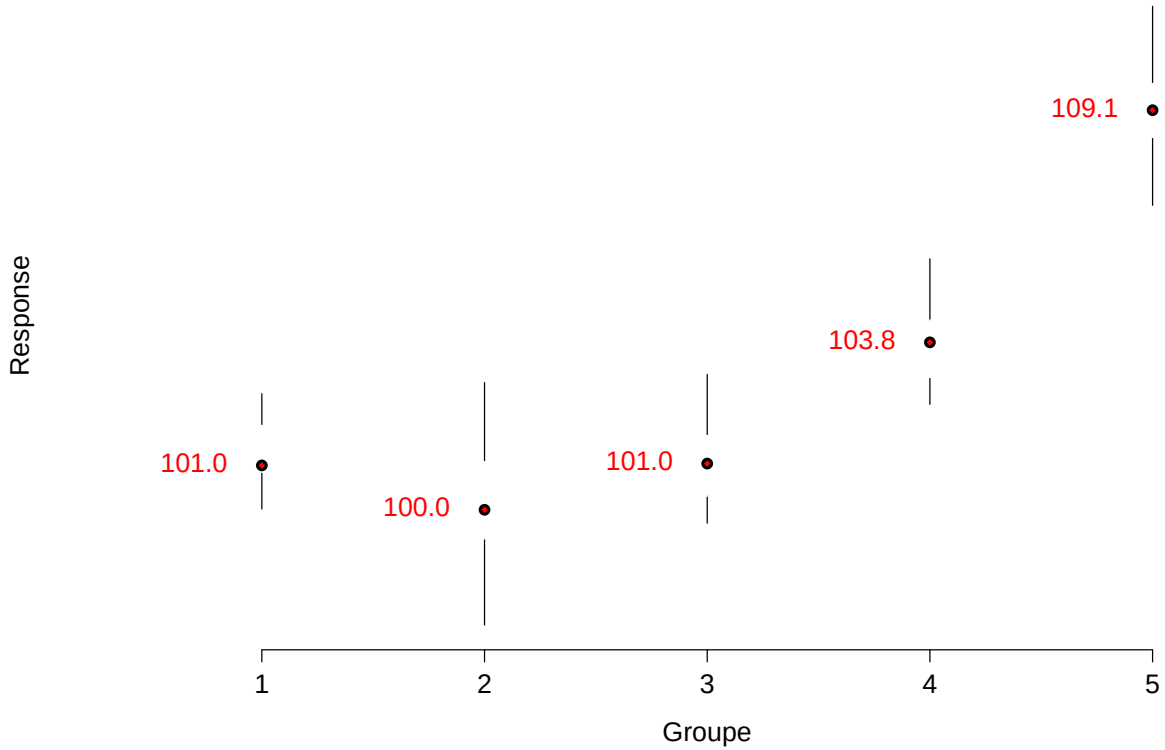


FIGURE – Distribution of a continuous variable on 4 groups

HAVE WE LOST SOMETHING ?

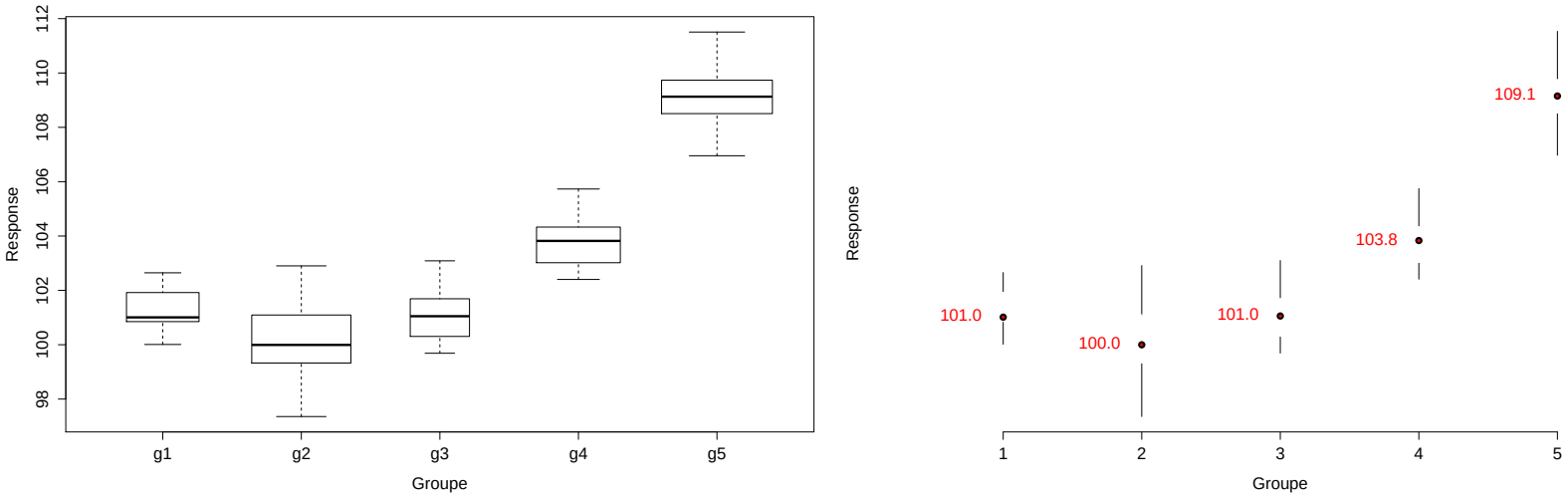


FIGURE – Distribution of a continuous variable on 4 groups

Did you noticed that group 1 and group 3 had the same median (101.0) ? see the *ggplot* theme `+ theme_tufte()`

Integrity : The lie factor

$$LieFactor = \frac{Size\ of\ effect\ shown\ in\ graphic}{Size\ of\ effect\ in\ data} \quad (1)$$

A Lie Factor $\neq 1$ indicates a substantial distortion

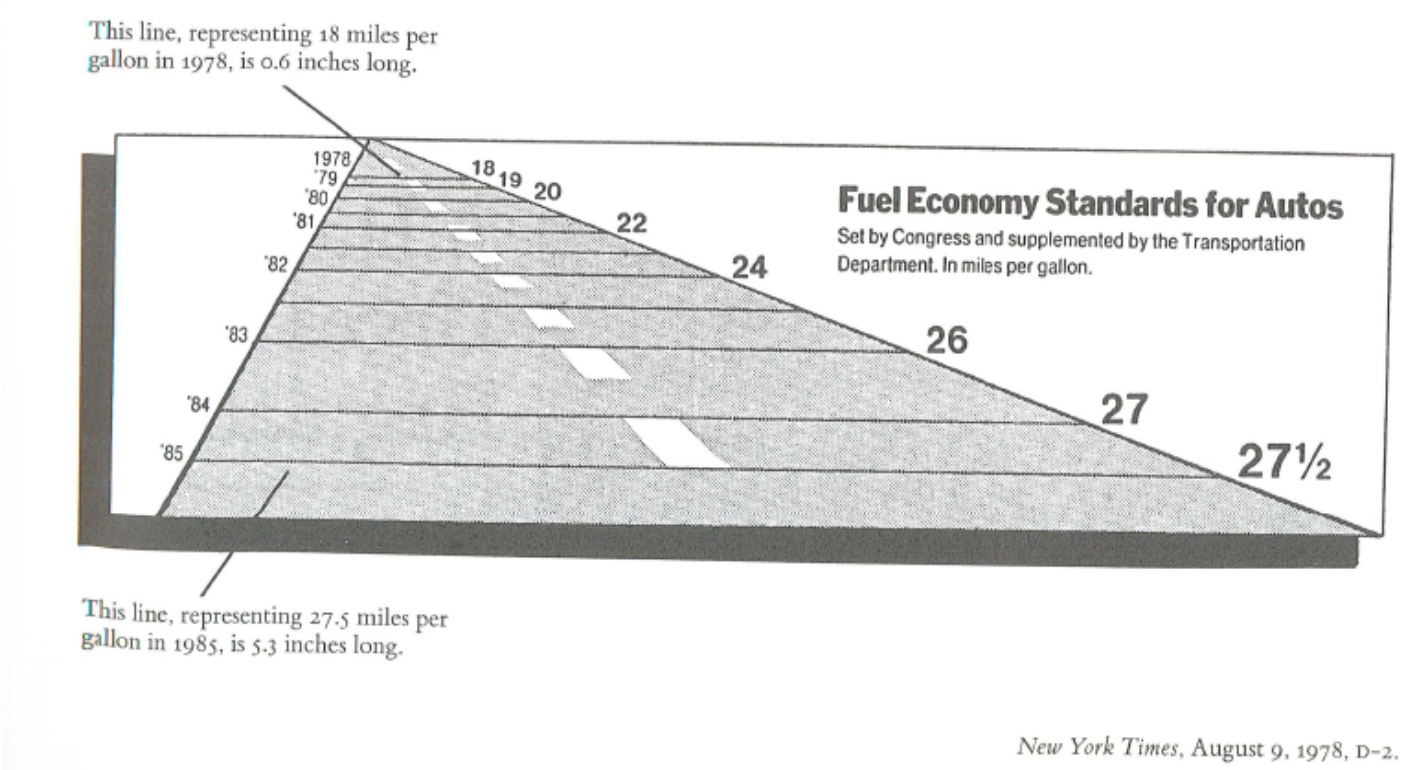


Figure – Fuel economy standards. (E. Tufte - from NY Times 1978)

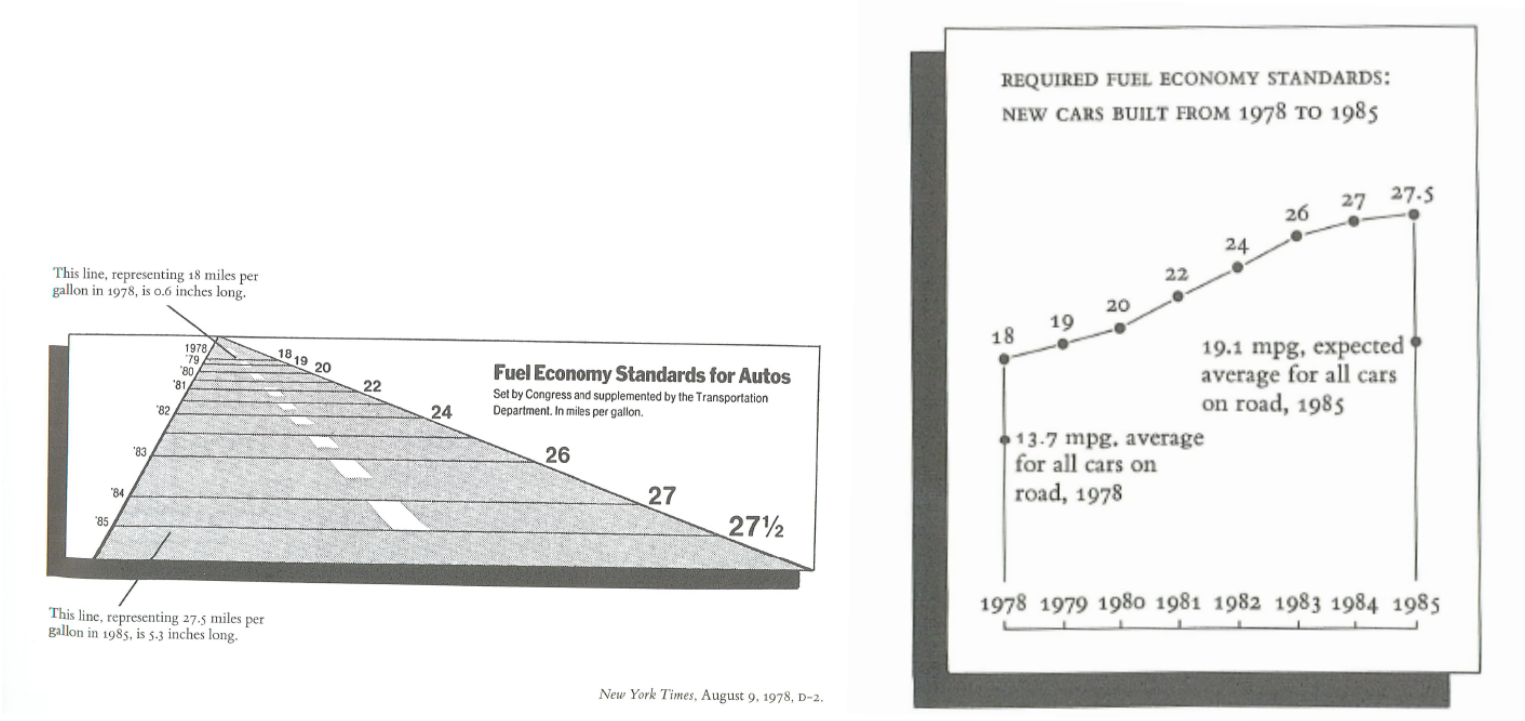


FIGURE – Fuel economy standards (revisited)

The "18 mpg" line measures 1.5 cm (in 1978) ; the "27,5 mpg" measures 13 cm (in 1985)

$\longrightarrow$ Lie factor = 14.5% ! ! !

Practical example : Context

In the next 5 minutes, you will design **"the best way"** to **compare** two numbers

75 and **37**.

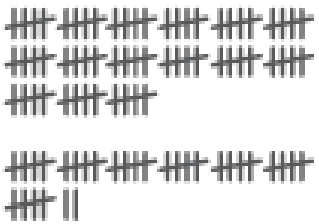
From S. Ortiz

CONTEXT MATTERS !

75, 37

a

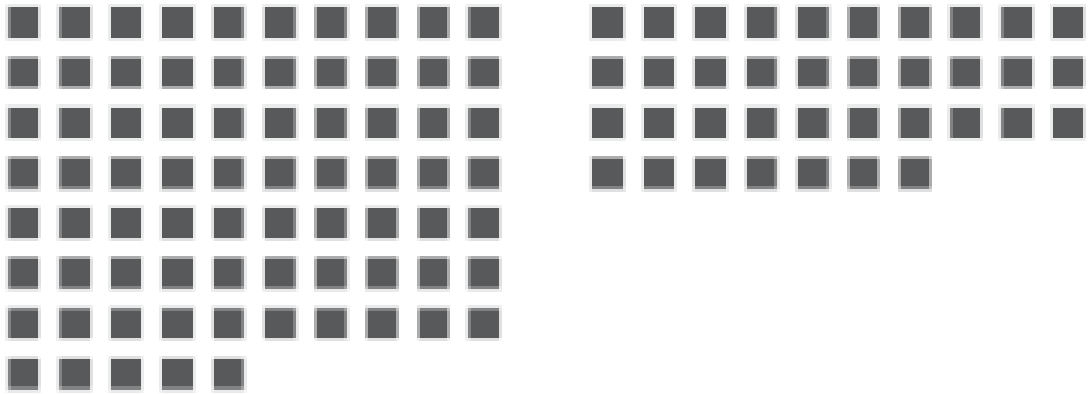
b



c

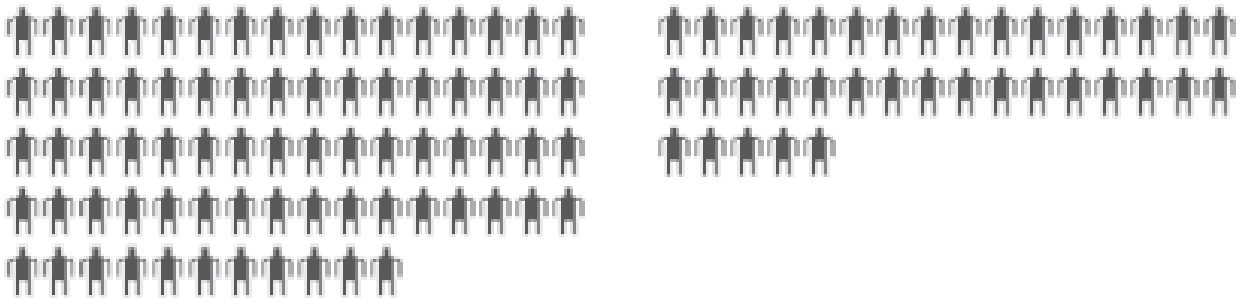
Solutions proposed by S. Ortiz

CONTEXT MATTERS !



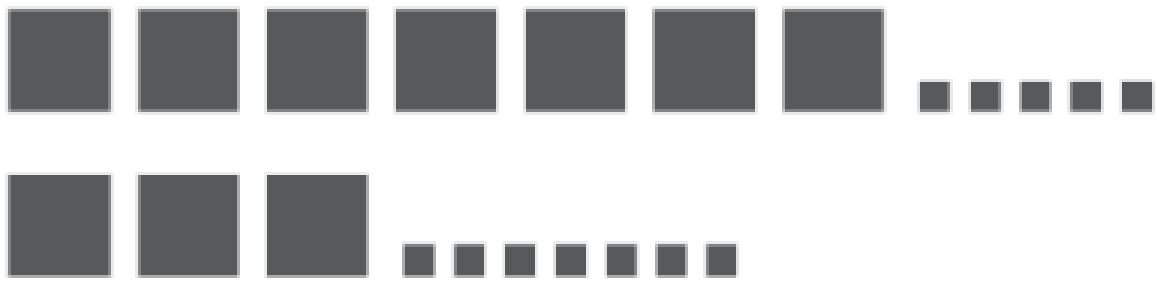
Solutions proposed by S. Ortiz

Context matters !



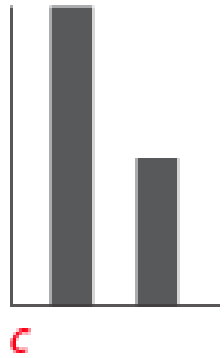
Solutions proposed by S. Ortiz

CONTEXT MATTERS !



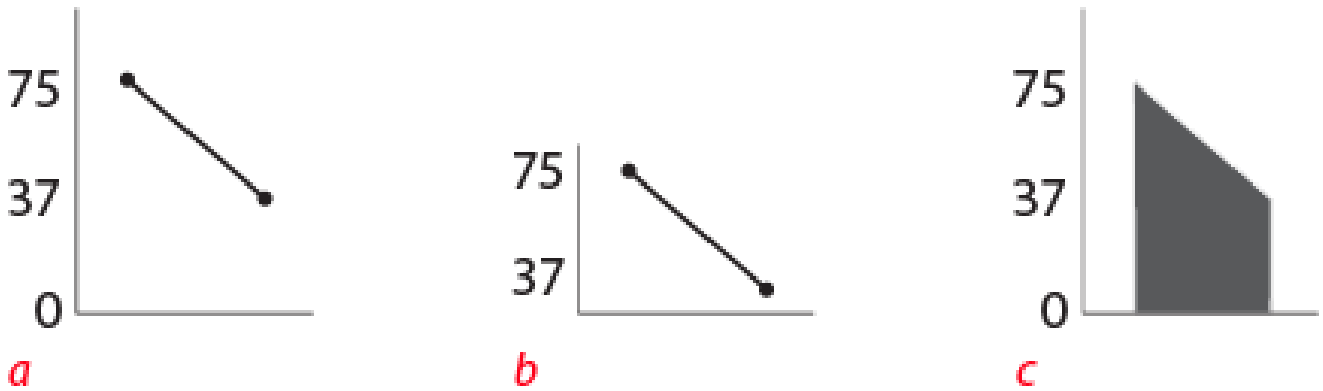
Solutions proposed by S. Ortiz

CONTEXT MATTERS !



Solutions proposed by S. Ortiz

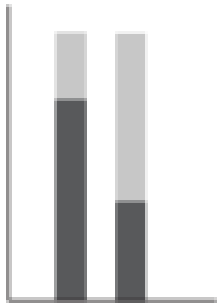
CONTEXT MATTERS !



Solutions proposed by S. Ortiz

CONTEXT MATTERS !

a



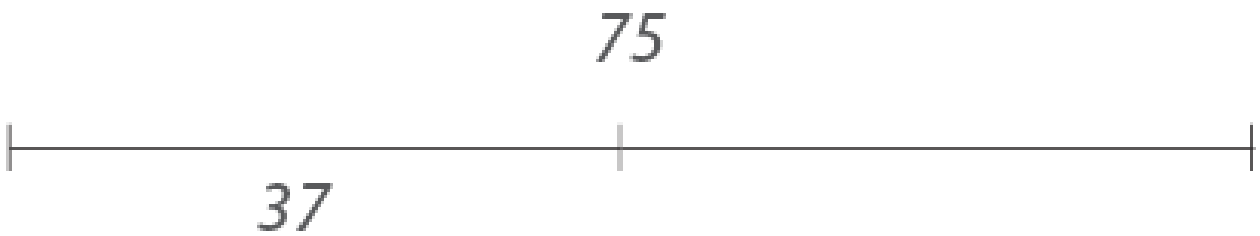
b

Solutions proposed by S. Ortiz

CONTEXT MATTERS !

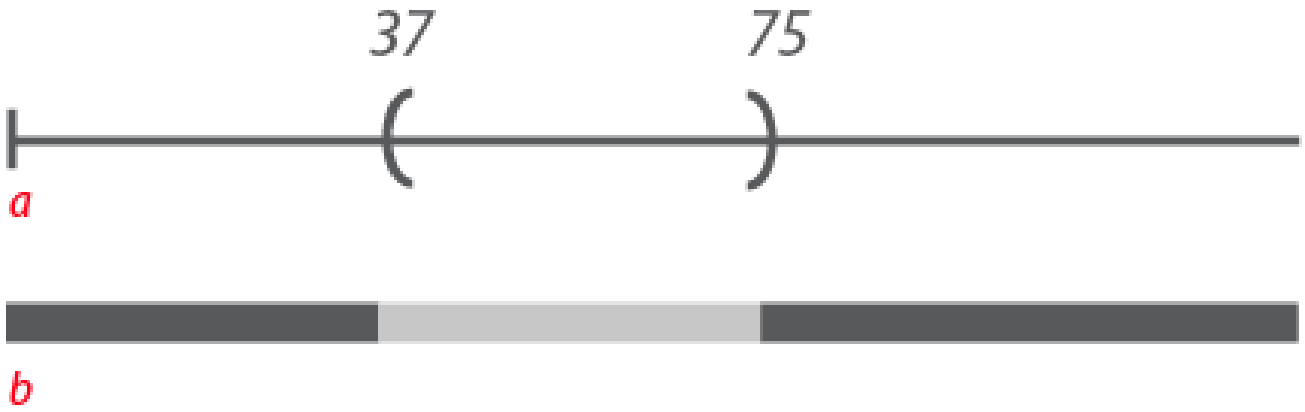
Solutions proposed by S. Ortiz

CONTEXT MATTERS !



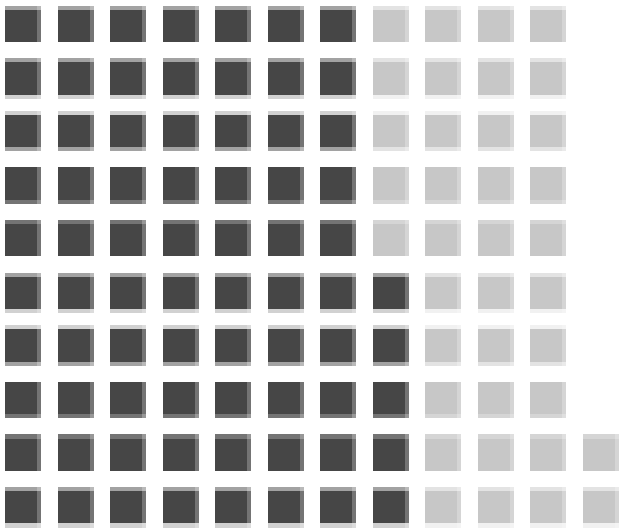
Solutions proposed by S. Ortiz

CONTEXT MATTERS !



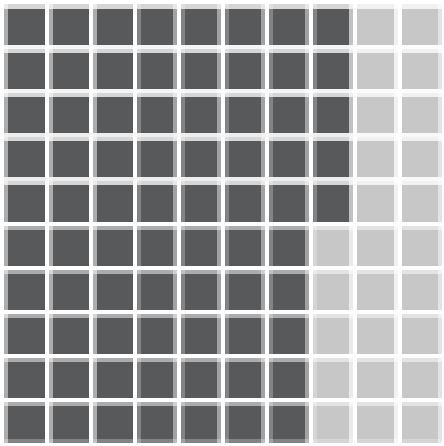
Solutions proposed by S. Ortiz

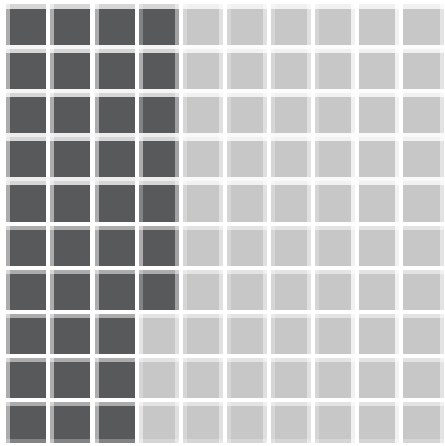
Context matters !



Solutions proposed by S. Ortiz

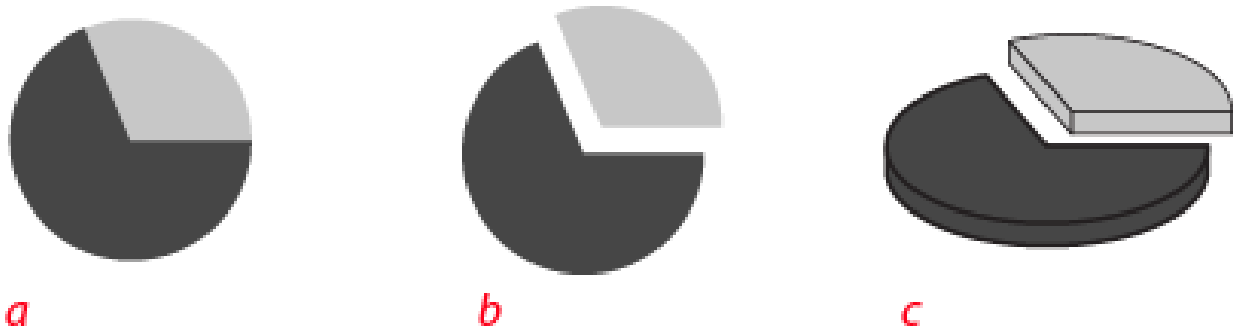
CONTEXT MATTERS !





Solutions proposed by S. Ortiz

CONTEXT MATTERS !



Solutions proposed by S. Ortiz

Context matters !

Solutions proposed by S. Ortiz

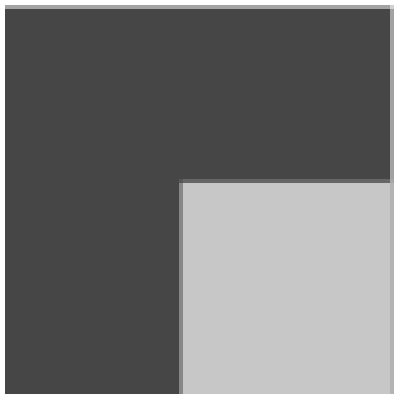
CONTEXT MATTERS !

Solutions proposed by S. Ortiz

Context matters !

Solutions proposed by S. Ortiz

Context matters !

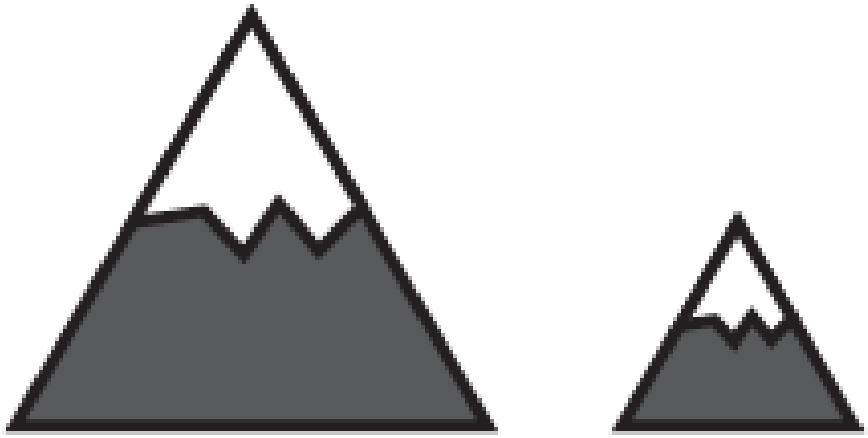


Solutions proposed by S. Ortiz

Context matters !

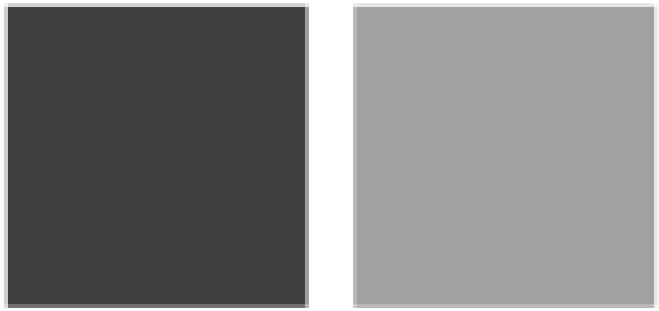
Solutions proposed by S. Ortiz

CONTEXT MATTERS !



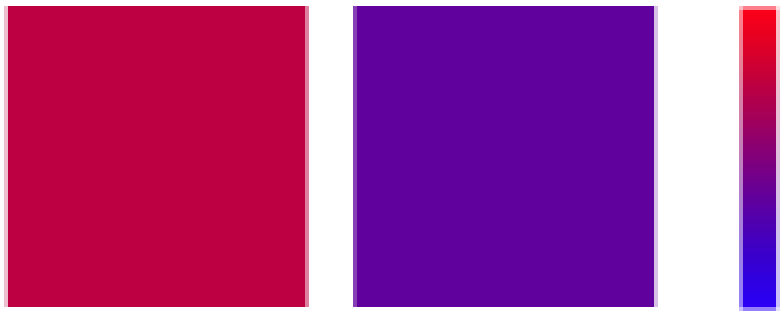
Solutions proposed by S. Ortiz

CONTEXT MATTERS !



Solutions proposed by S. Ortiz

Context matters !



Solutions proposed by S. Ortiz

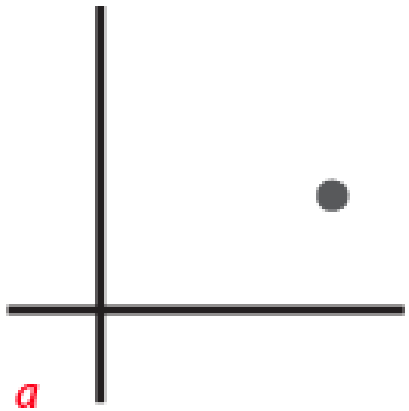
CONTEXT MATTERS !

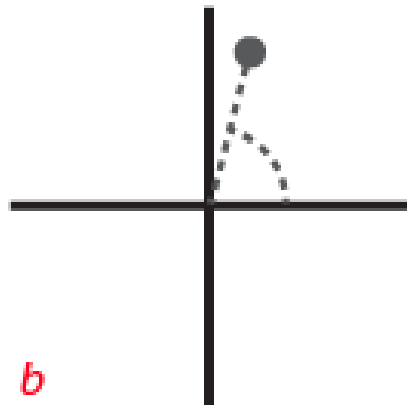
Solutions proposed by S. Ortiz

CONTEXT MATTERS !

Solutions proposed by S. Ortiz

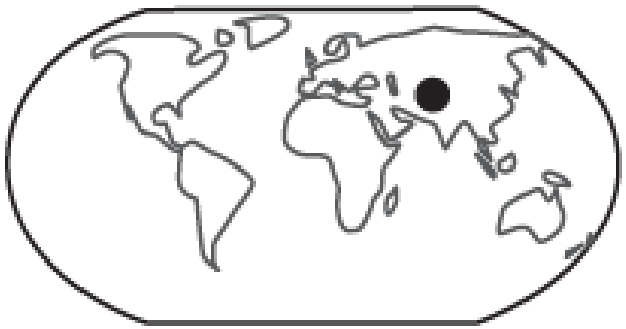
CONTEXT MATTERS !



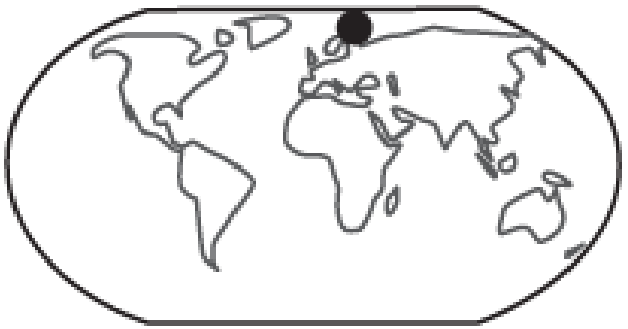


Solutions proposed by S. Ortiz

CONTEXT MATTERS !



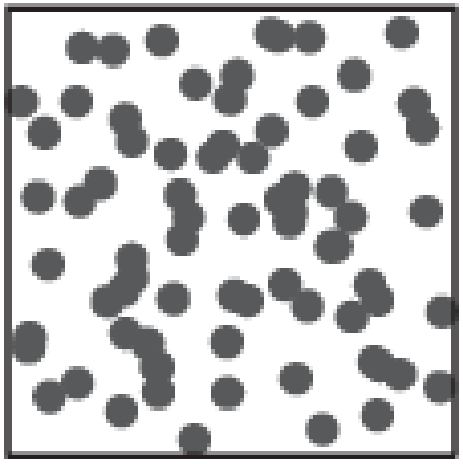
a

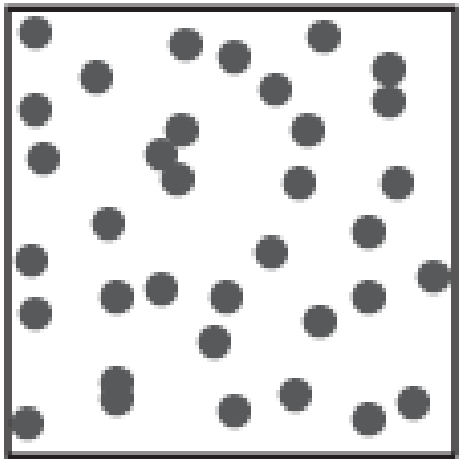


b

Solutions proposed by S. Ortiz

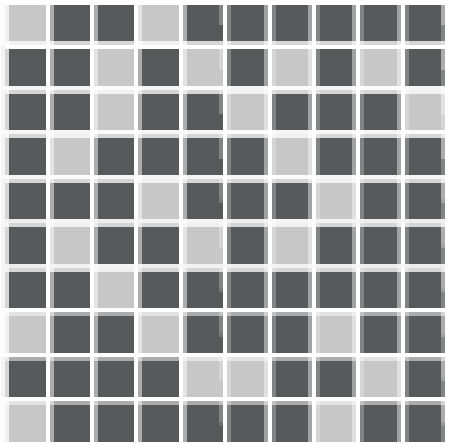
CONTEXT MATTERS !

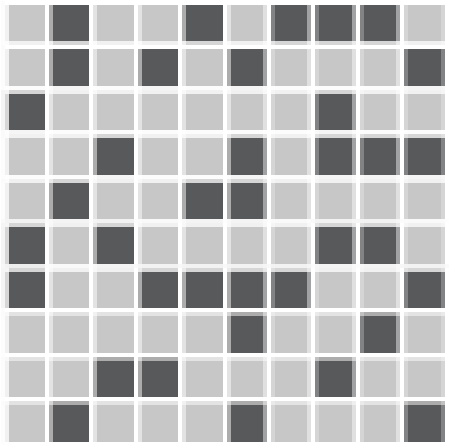




Solutions proposed by S. Ortiz

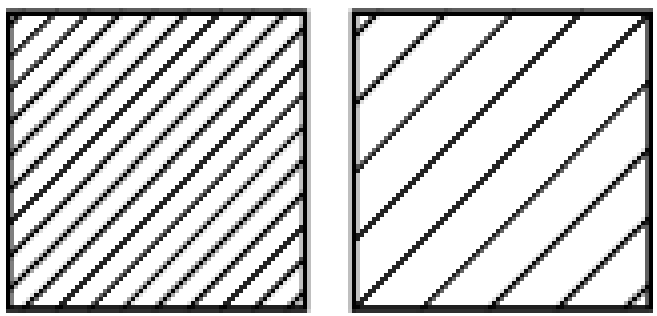
CONTEXT MATTERS !





Solutions proposed by S. Ortiz

Context matters !



Solutions proposed by S. Ortiz

CONTEXT MATTERS !

Solutions proposed by S. Ortiz

Context matters (follow-up)

What are the problems you've encountered ?

Context matters (follow-up)

What are the problems you've encountered ?

- Context ?

Context matters (follow-up)

What are the problems you've encountered ?

- Context ?
- Audience ?

Context matters (follow-up)

What are the problems you've encountered ?

- Context ?
- Audience ?
- What to compare ?

Context matters (follow-up)

What are the problems you've encountered ?

- Context ?
- Audience ?
- What to compare ?
- Units ?

Context matters (follow-up)

What are the problems you've encountered ?

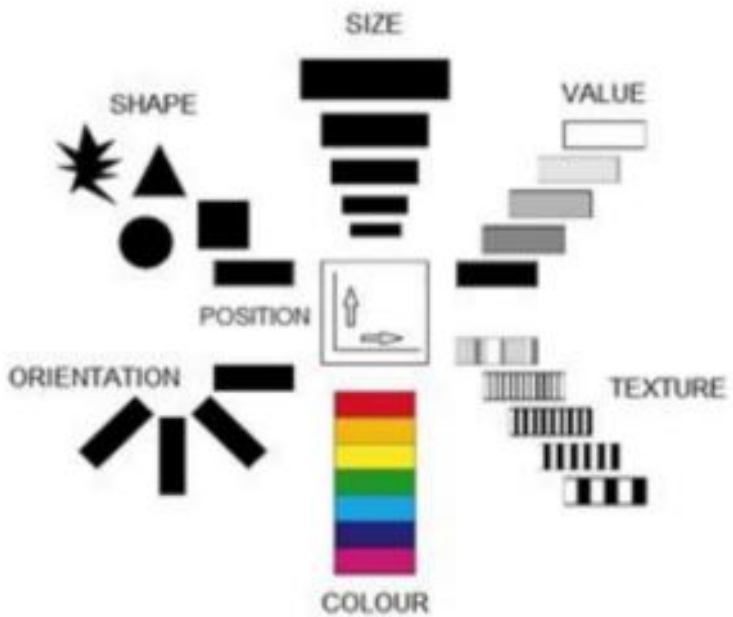
- Context ?
- Audience ?
- What to compare ?
- Units ?
- Multiple solutions ?

[Rules :] Bertin's approach (A visual language)

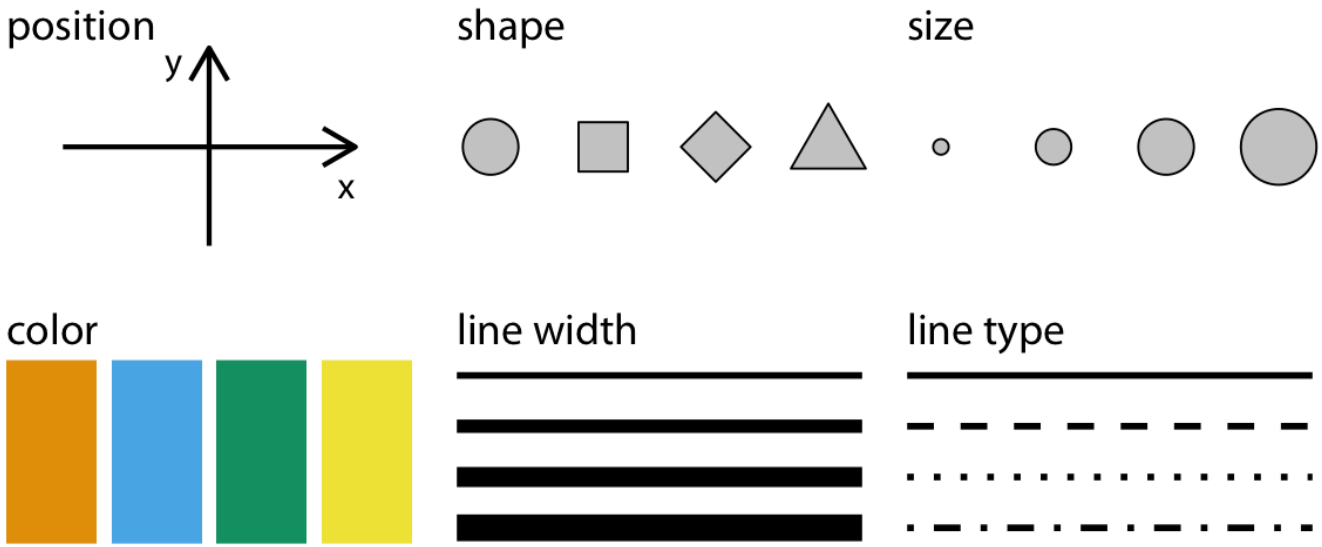
If graphs are used to communicate, it is a form of language. Any language has a grammar, "words" and logic. Let us study the science that deals with signs or sign language : *"The Semiology"*.

Table – Bertin's definition of 8 visual variables

Position (x, y)
Size
Value
Texture
Colour
Orientation
Shape



A VISUAL LANGUAGE



From *Claus O. Wilke*

THESE VARIABLES SERVE DIFFERENT GOALS

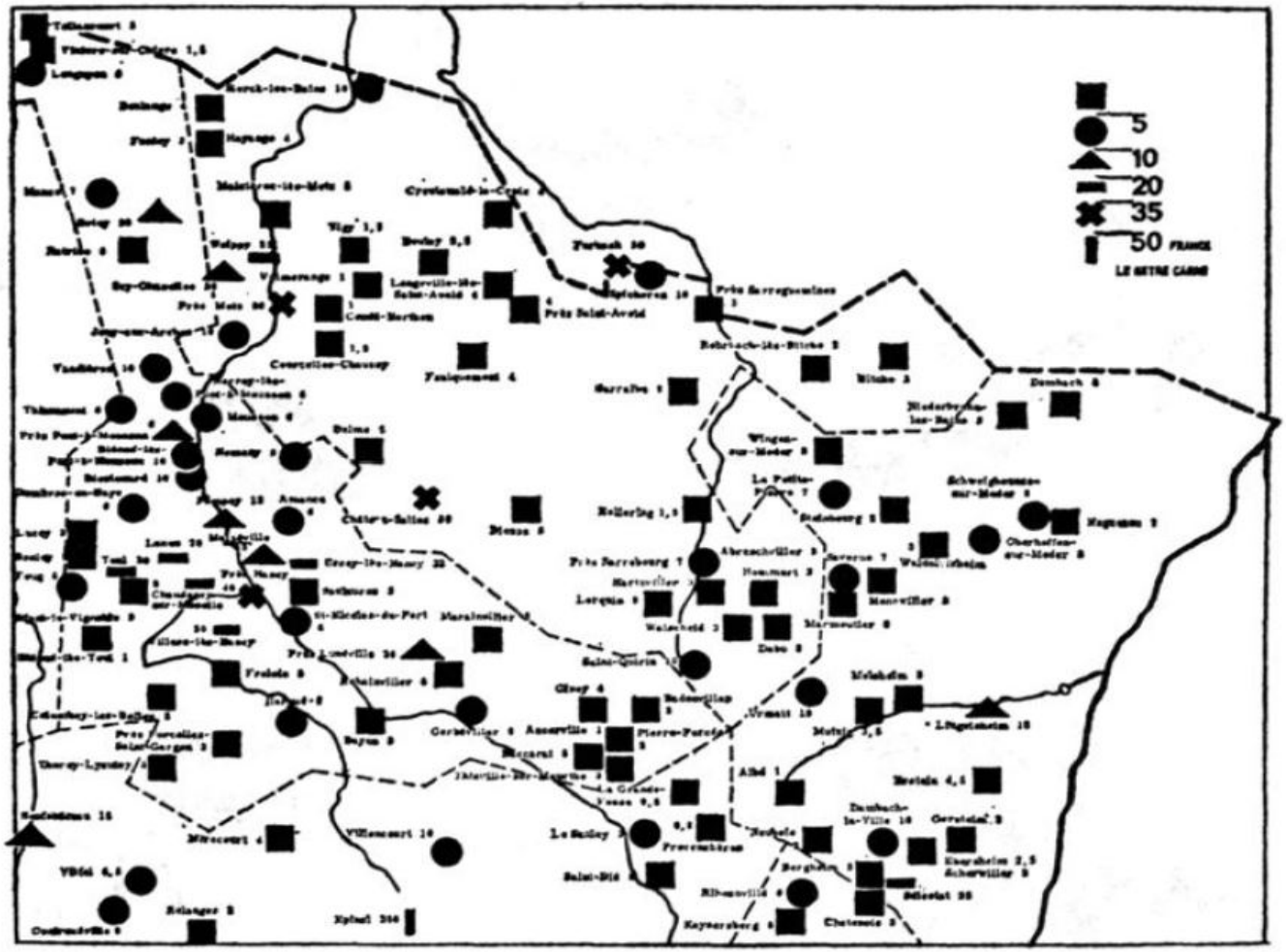
Visual variable *syntactics*, designating each visual variable as suited or not for levels of measurement :
Equivalence, differences, order, proportions.

	Visual variable syntactics :			
	Variable suited for :			
Position (x, y)	$\equiv$	$\neq$	O	$\propto$
Size	$\equiv$	$\neq$	O	$\propto$
Value	$\equiv$	$\neq$	O	$\propto$
Texture	$\equiv$	$\neq$	O	
Colour	$\equiv$	$\neq$		
Orientation	$\equiv$	$\neq$		
Shape	$\equiv$			

$\equiv$: Equivalence $\neq$: Differences ; O : Order ; $\propto$: Proportions

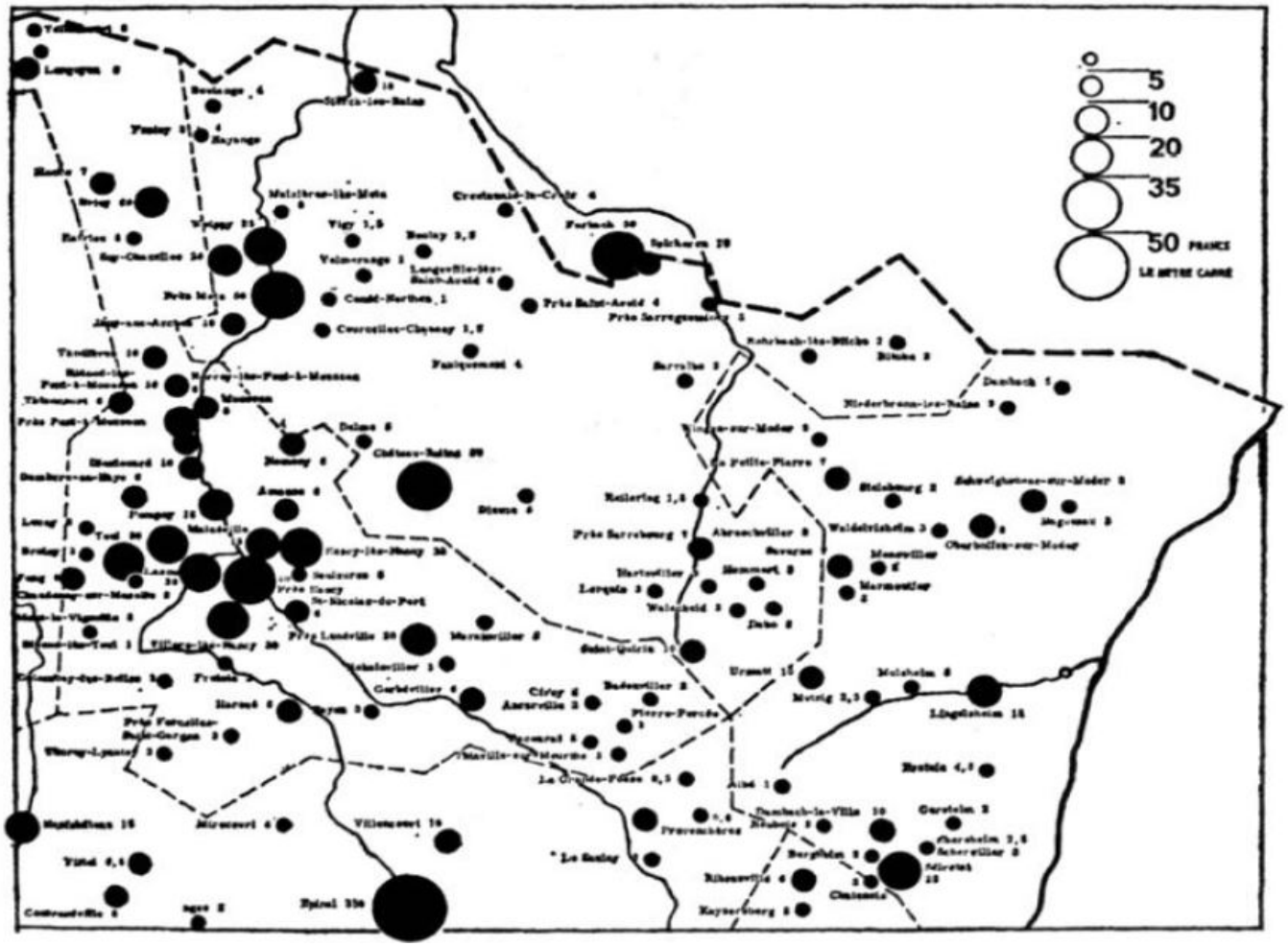
SHAPE IS NOT SUITABLE FOR PROPORTIONALITY

Price of land in the East of France Bertin (1970)



SIZE IS SUITABLE FOR PROPORTIONALITY

Price of land in the East of France Bertin (1970)

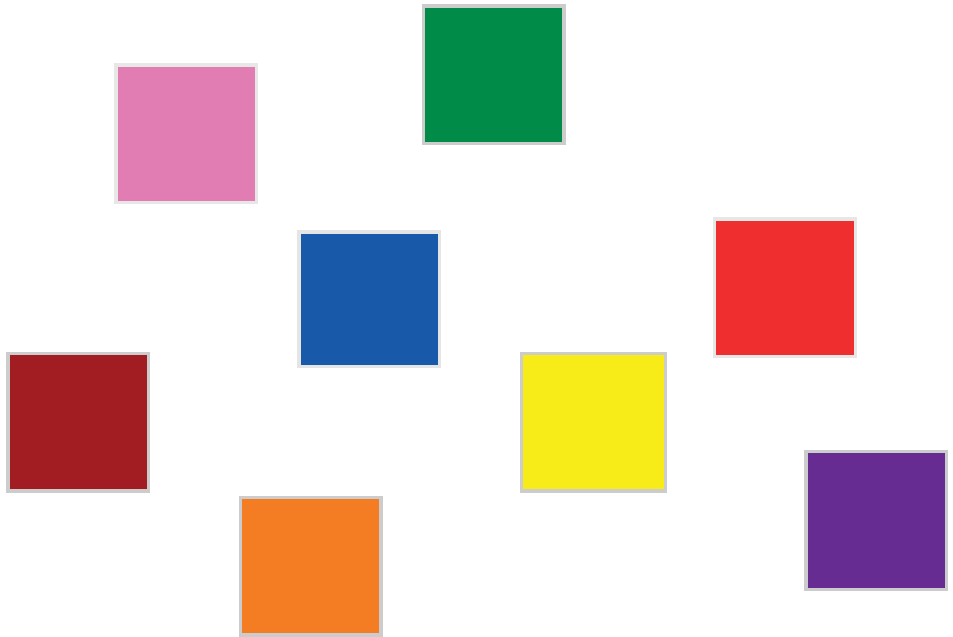


WHAT ELSE ?

Is color suitable for proportionality ?

A NOTE ON COLORS

"Colors" are not suited for ordering nor for proportionality !
Try putting the following hues in order from low to high.



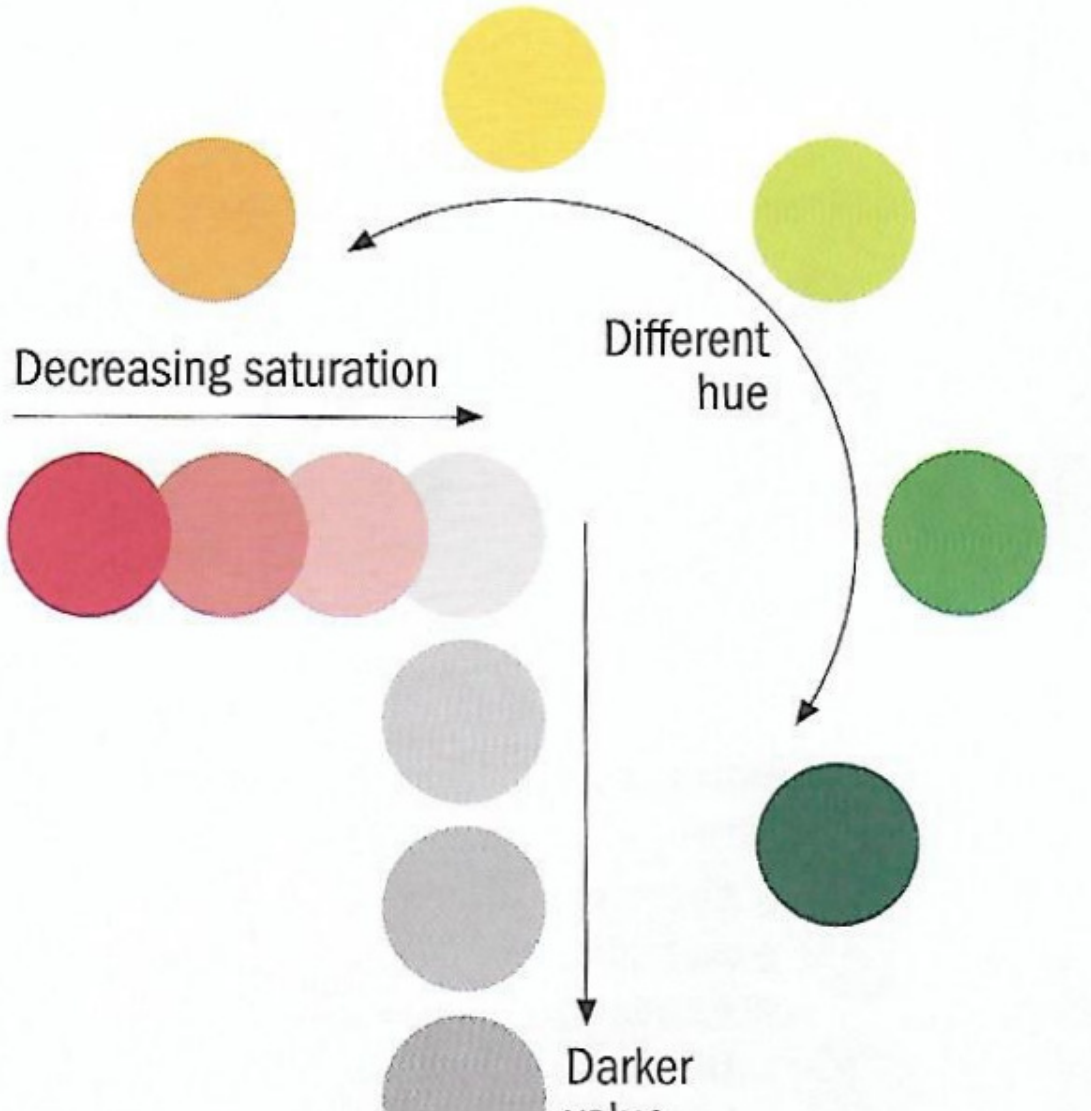
A NOTE ON COLORS

These colors are easy to order from low to high.
Few (2008) provides meaningful solutions for choosing palettes of colors, for example for heatmaps.

See also the *ggplot* theme `theme_few()`

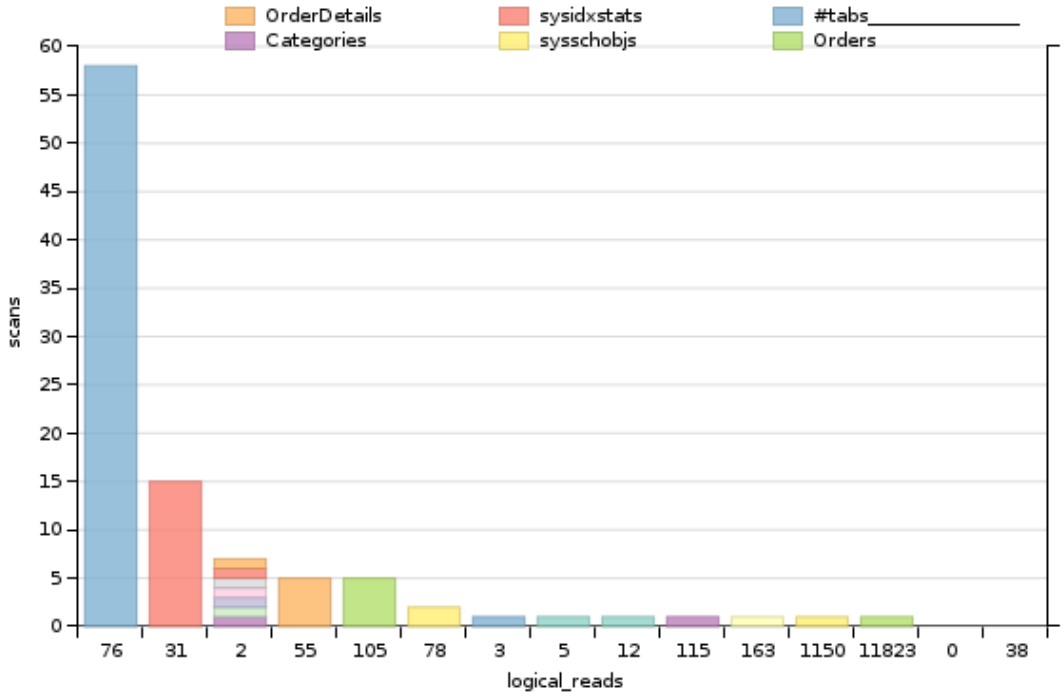
A NOTE ON COLORS

What we call "colors" are three dimensional objects



A NOTE ON COLORS (FINAL)

Colors are sometimes a graphic puzzle Tufte (2001).



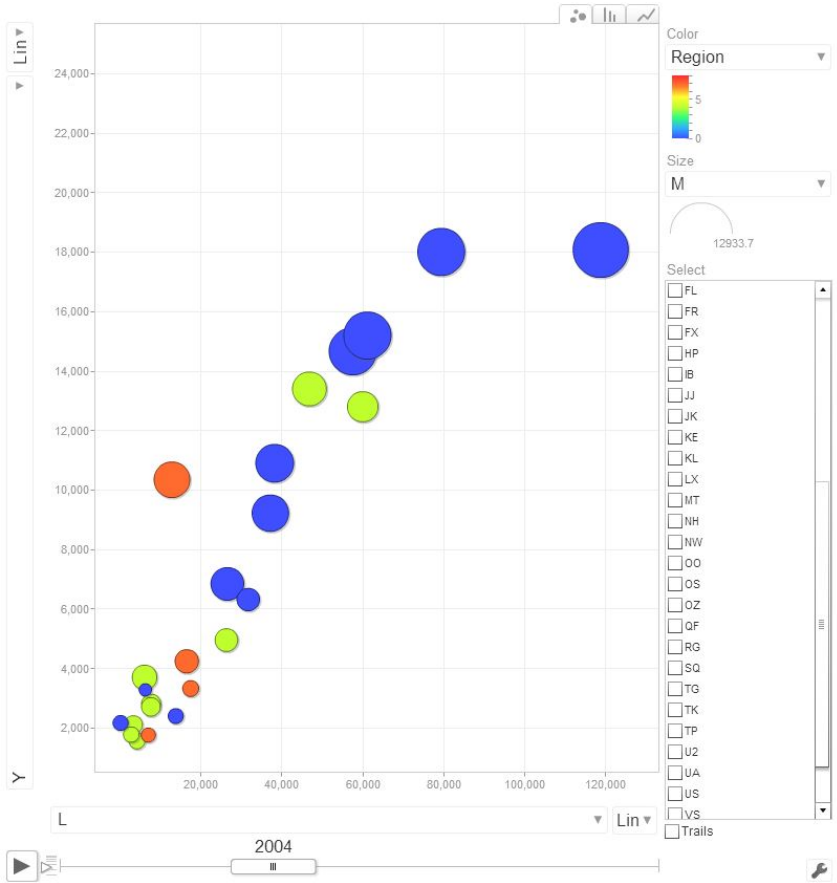
scans

Your eyes will go back and forth from the graph to the legend...

Source `http://viz.wtf/image/135265269618`)

Conjunction of colours and proportionality

Productivity of Airlines



(Demo with googleVis)

A QUIZ ON PROPORTIONALITY :

If 100% of the US prisoners are represented by the big green square...what is the percentage for each group ?

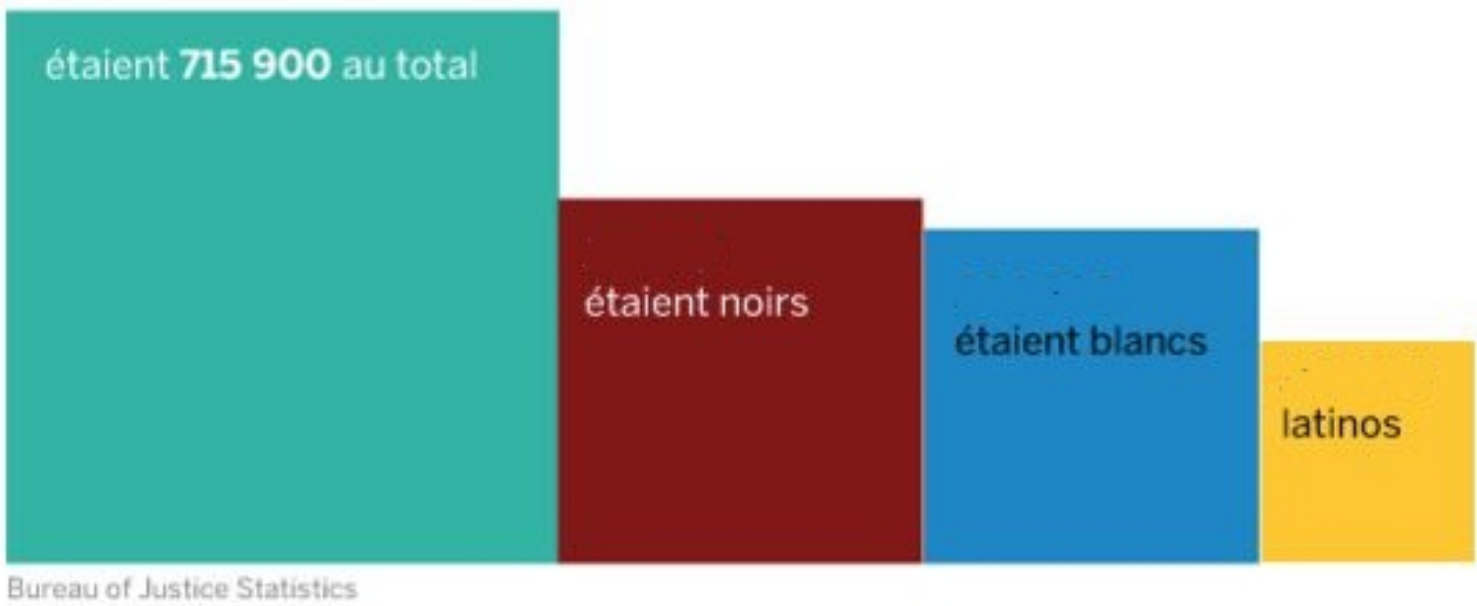


FIGURE – Ethic composition of prisoners in Jail in 2008 in the USA. *(Le Monde 5/12/2014)*

NOT SO SIMPLE…

If 100% of the US prisoners are represented by the big square...what is the percentage for each group ?

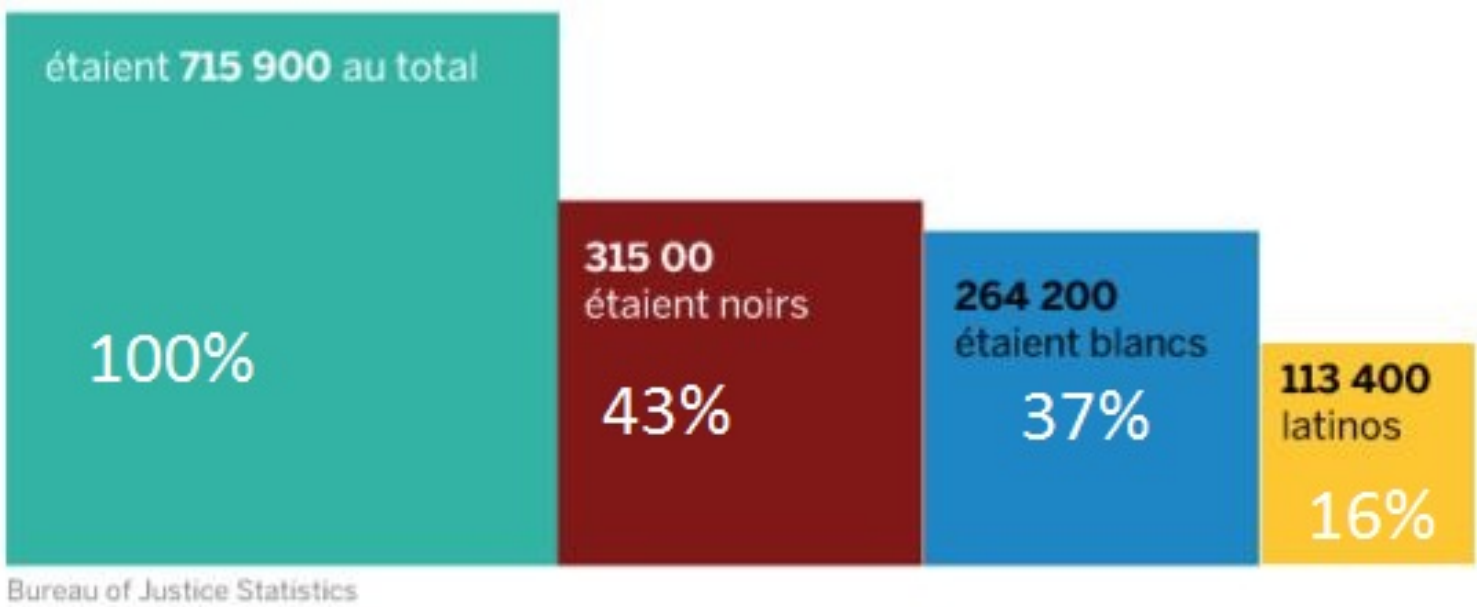
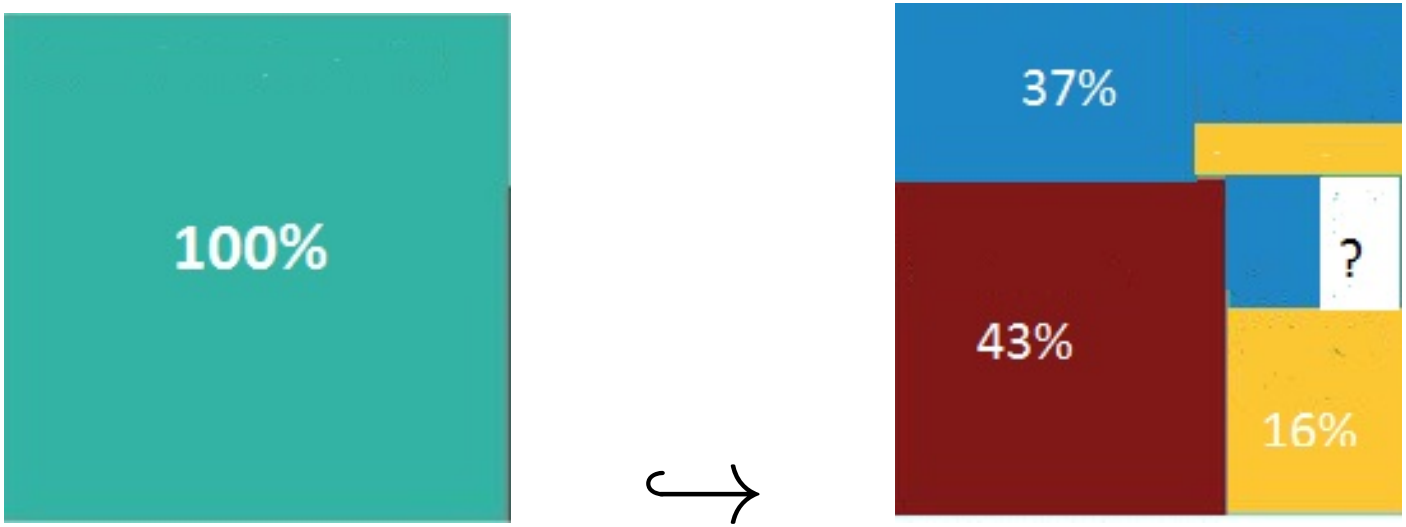


FIGURE – Ethic composition of prisoners in Jail in 2008 in the USA. (Le Monde 5/12/2014)

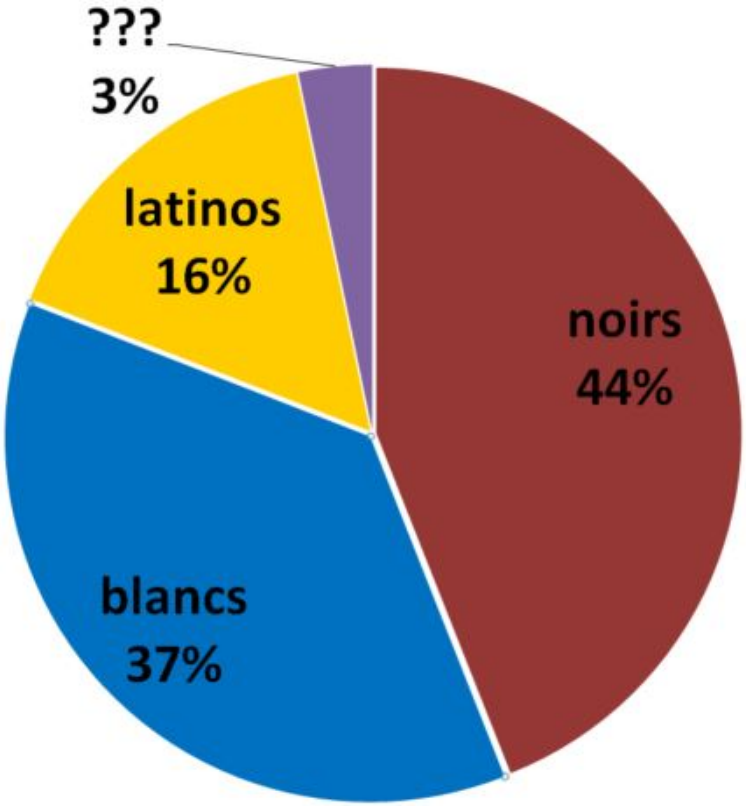
VERIFICATION

If 100% of the US prisoners are represented by the big square...what is the percentage for each group ?



Or...

If 100% of the US prisoners are represented by the big square...what is the percentage for each group ?



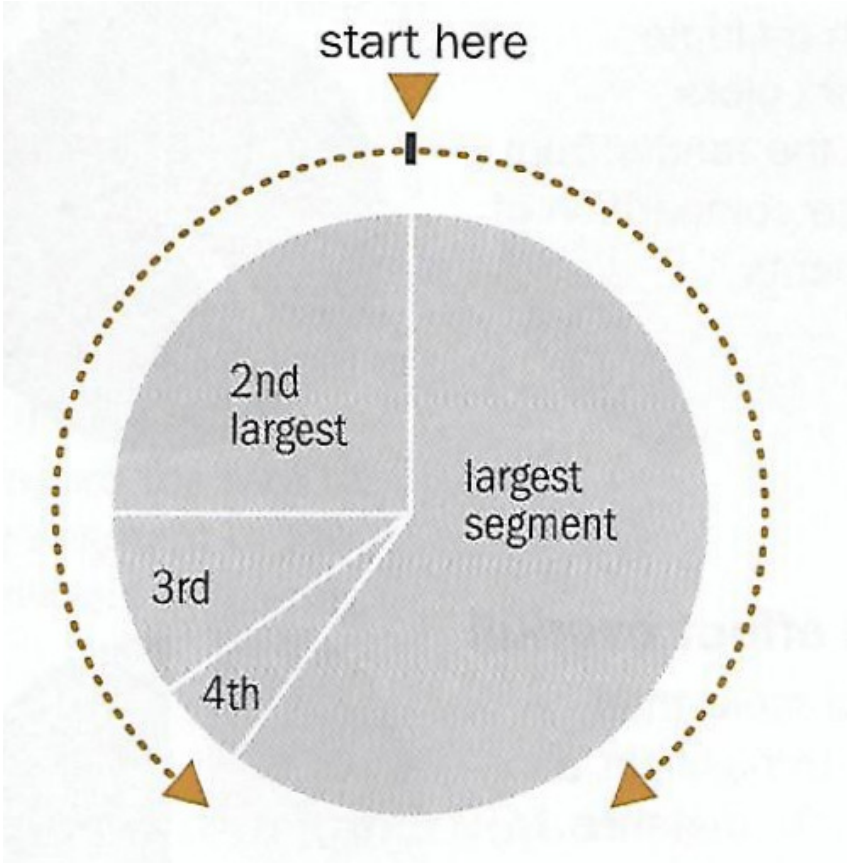
A note on Pie charts

Pie charts are graphics that can be used if :

- Few categories
- "good" distribution of shares
- "good" choice of colors

A NOTE ON PIE CHARTS (CONTINUED)

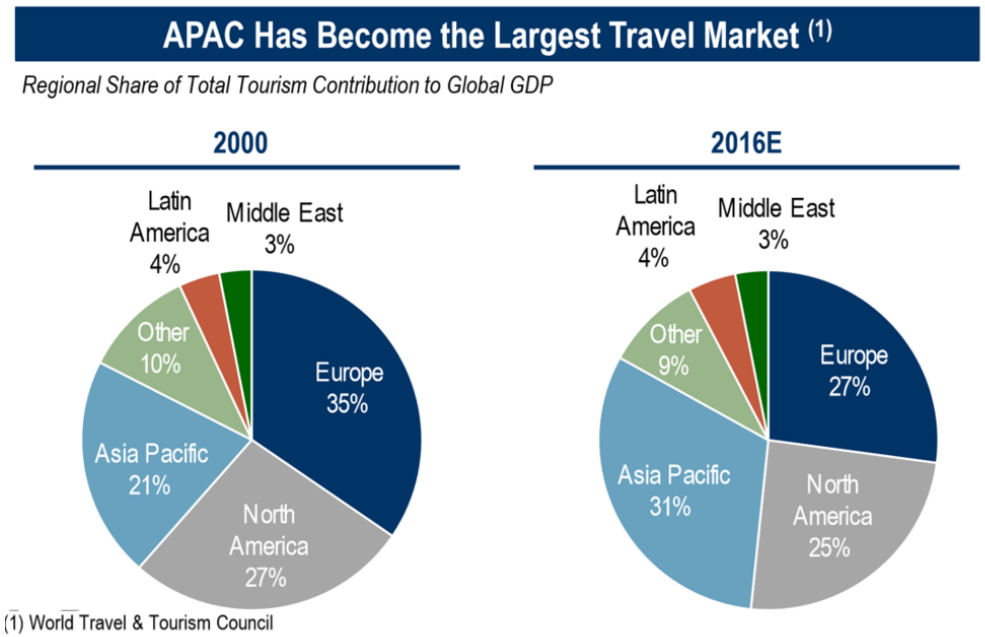
One should cleverly use of the reference line



from Wong (2010)

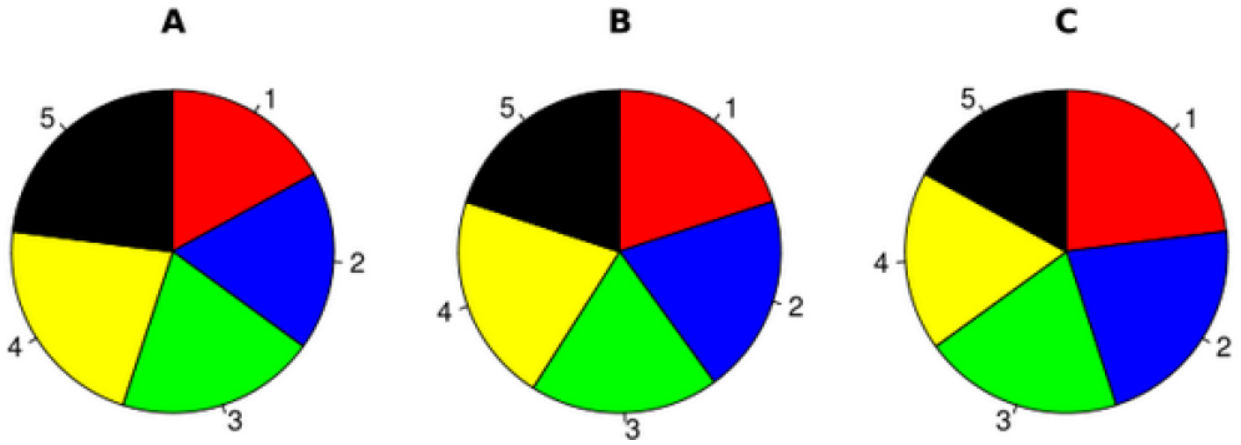
A NOTE ON PIE CHARTS (FINAL)

One should **not** use pie chart for comparisons

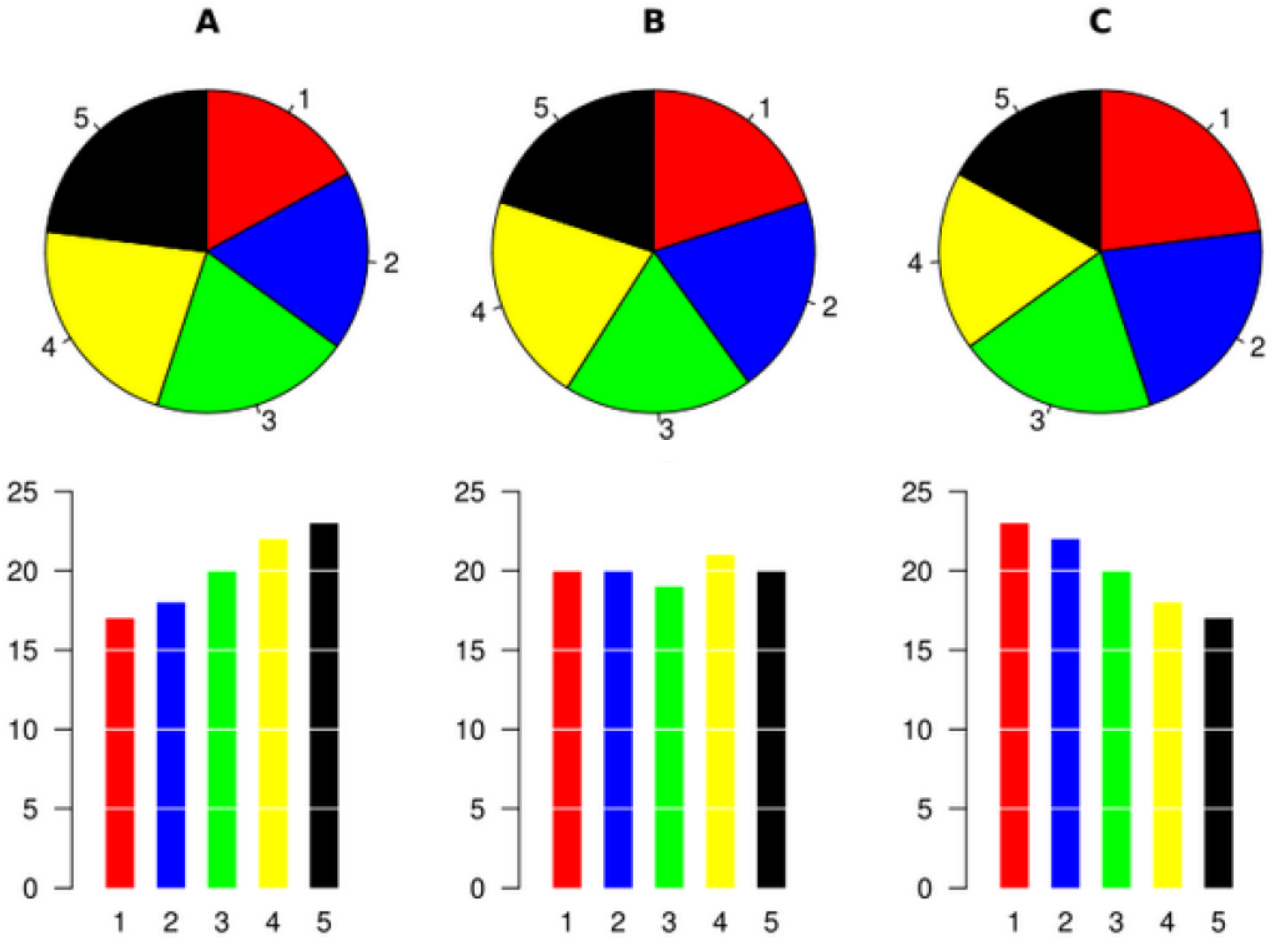


see SWD challenge

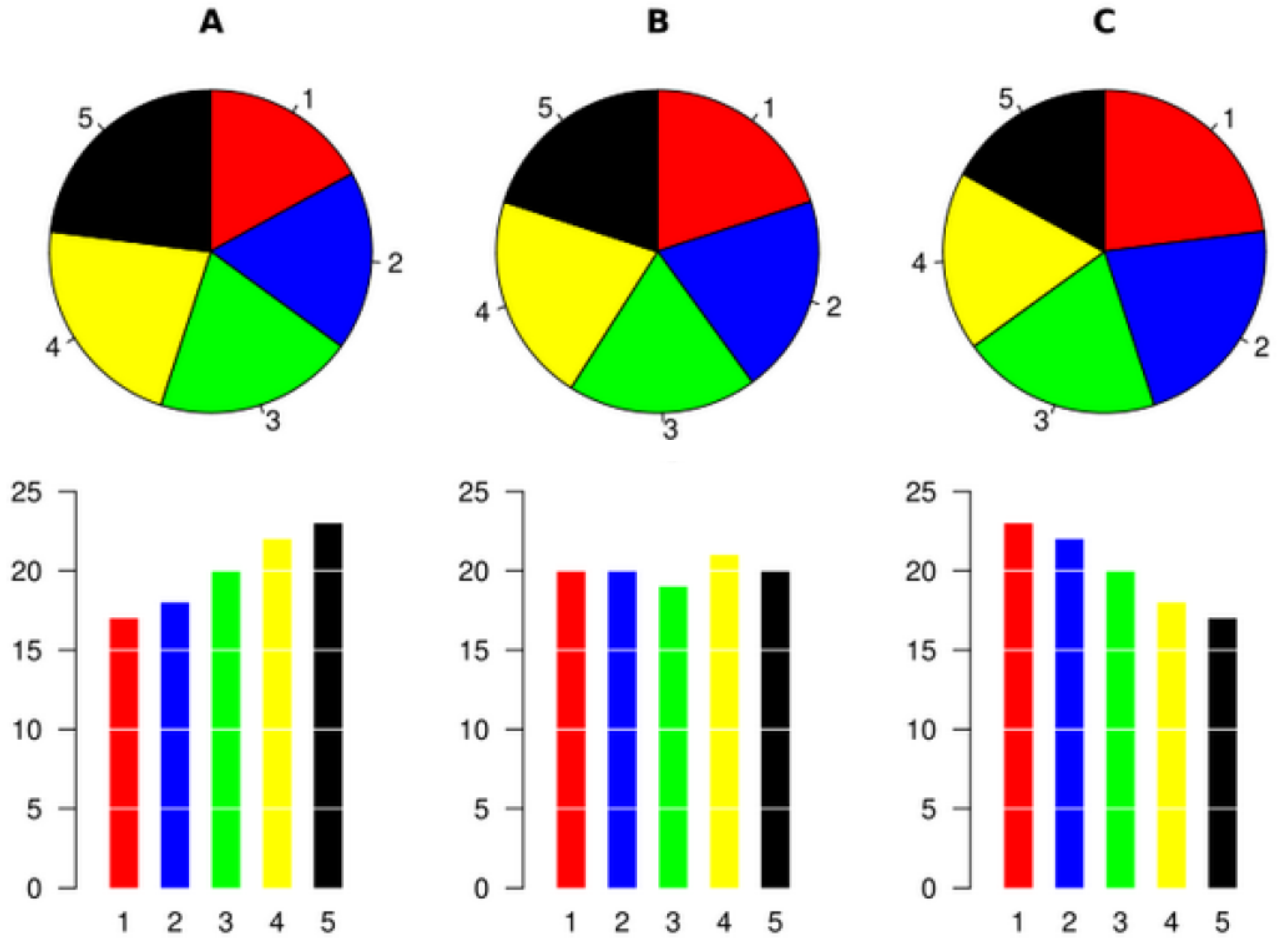
Be careful with Pie charts !



Be careful with Pie charts !



Be careful with Pie charts !



Source `https://twitter.com/freakonometrics/status/612742330160951296`

"Good" or "bad" Graphics ?

"There are no "good" nor "bad" graphics (...), there are graphics answering legitimate questions and graphics that do not answer question at all "

Bertin (1981)

SCHWABISH (JEP, 2014) BEFORE-AFTER

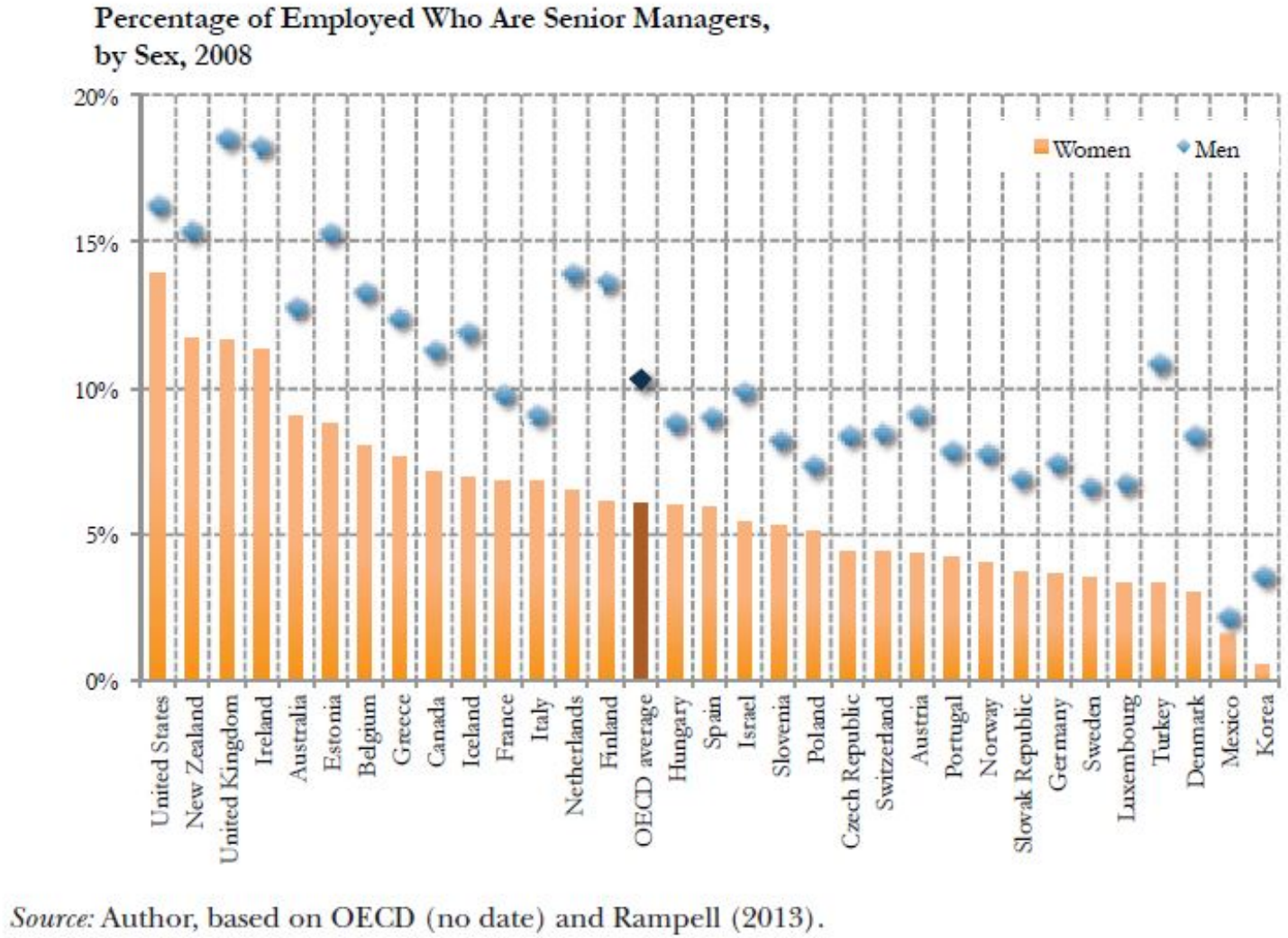
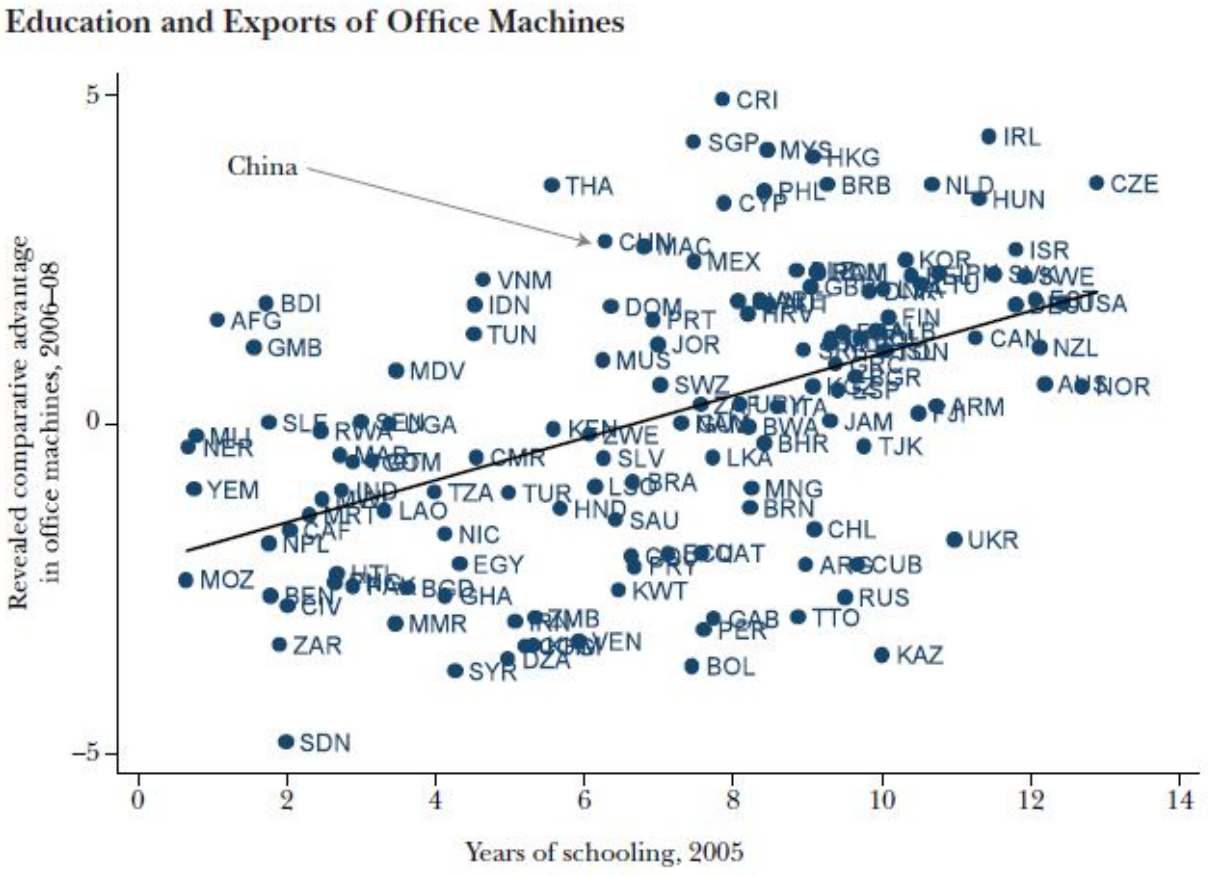


FIGURE – An Unbalanced Chart - **Original**

SCHWABISH (JEP, 2014) BEFORE-AFTER

FIGURE – An Unbalanced Chart - **Revised**

SCHWABISH (JEP, 2014) BEFORE-AFTER



Source: Hanson (2012).

FIGURE – A Clutterplot Example - **Original**

SCHWABISH (JEP, 2014) BEFORE-AFTER

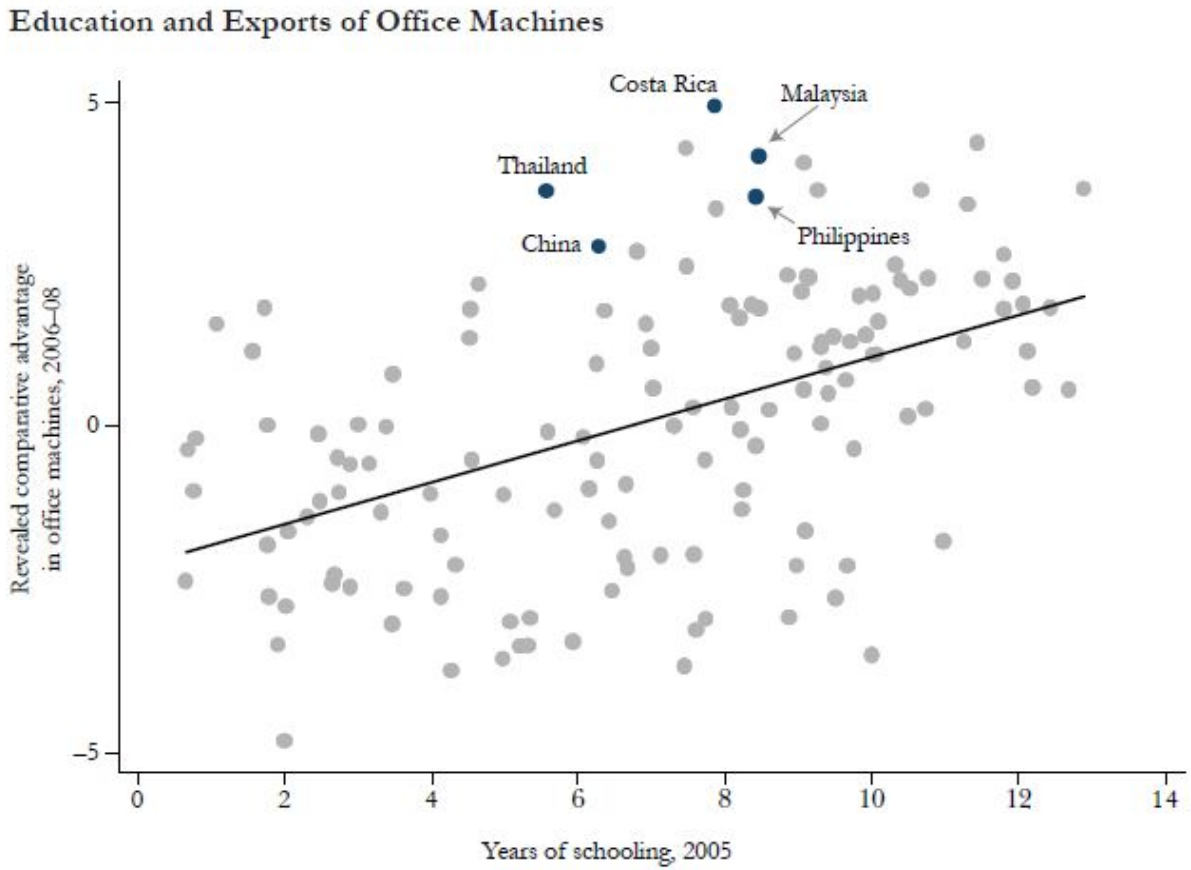
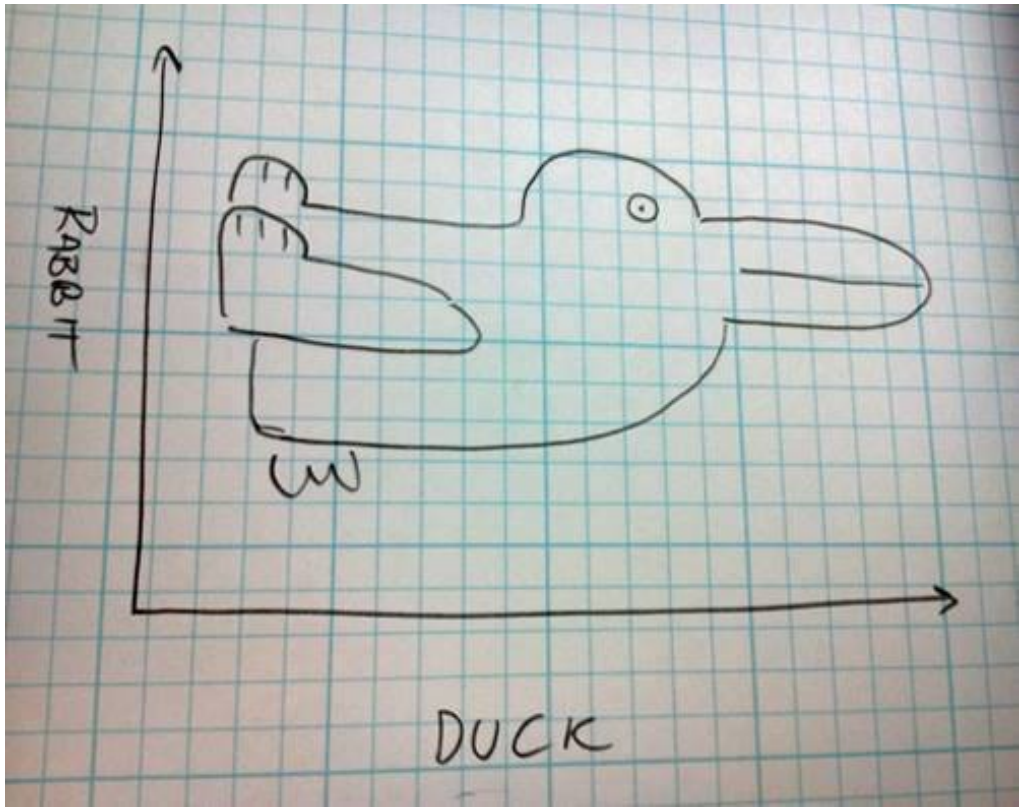


FIGURE – A Clutterplot Example - **Revised**

A NOTE ON PERCEPTION

A bird (Duck, Toucan ?) on the X axis, a rabbit on the Y axis !



//

Source

`http://flowingdata.com/2014/06/25/duck-vs-rabbit-plot/`

"PREATTENTIVE" VARIABLES

How many "3" in that sequence ? (from Ware (2012))

"PREATTENTIVE" VARIABLES

How many "3" in that sequence ? (from Ware (2012))

45929078059772098775972655665110049836645
27107462144654207079014738109743897010971
43907097349266847858715819048630901889074
25747072354745666142018774072849875310665

"PREATTENTIVE" VARIABLES

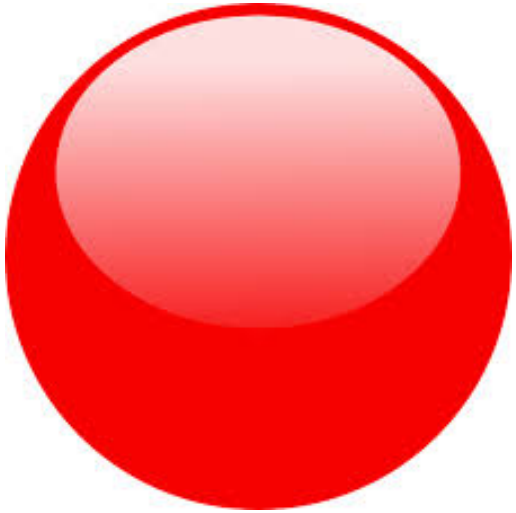
How many "3" in that sequence ? (from Ware (2012))

45929078059772098775972655665110049836645
27107462144654207079014738109743897010971
43907097349266847858715819048630901889074
25747072354745666142018774072849875310665

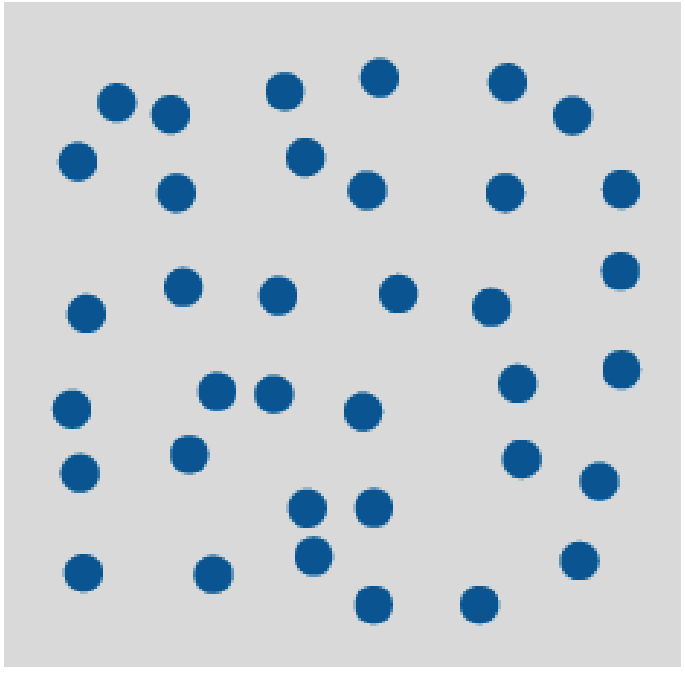
45929078059772098775972655665110049836645
27107462144654207079014738109743897010971
43907097349266847858715819048630901889074
25747072354745666142018774072849875310665

And now...

Find the red dot !

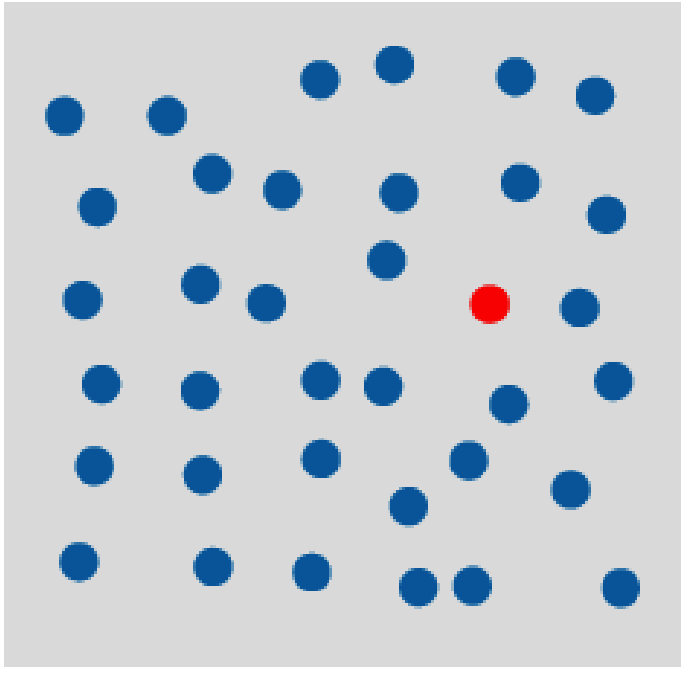


TEST : FIND THE RED DOT !



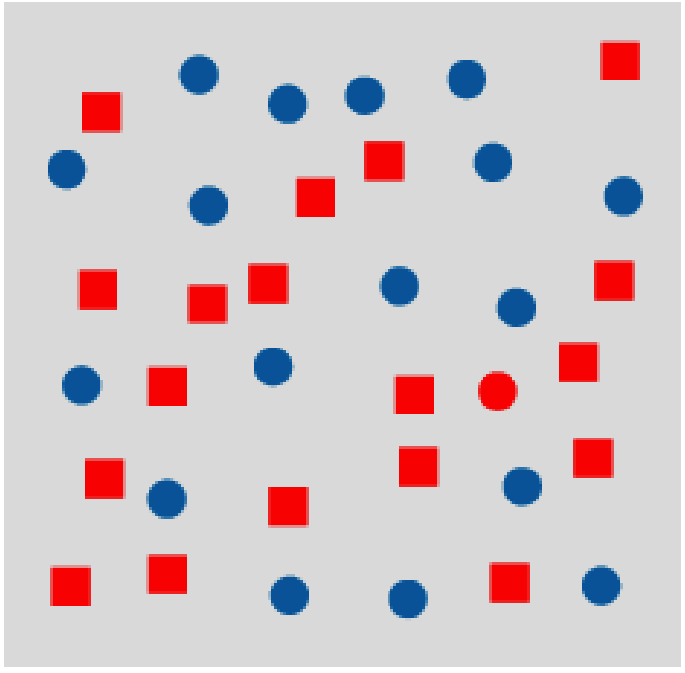
TEST : FIND THE RED DOT !

TEST : FIND THE RED DOT !



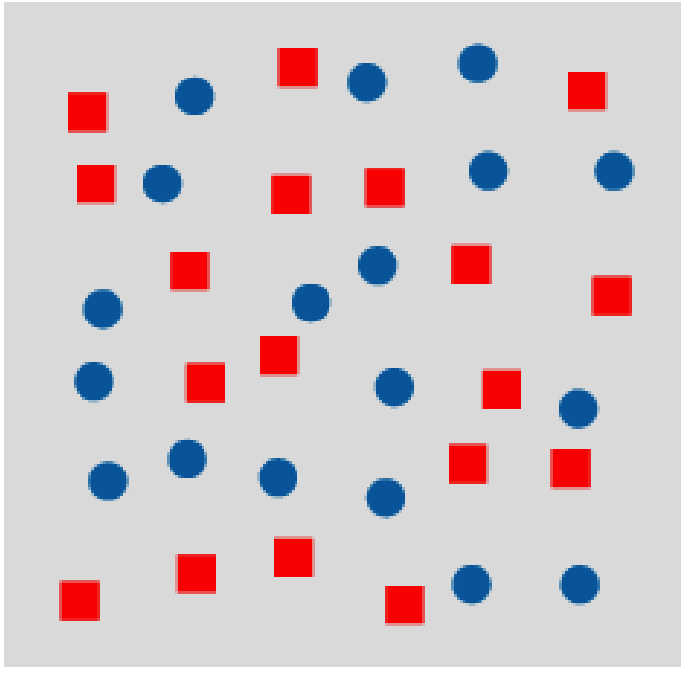
TEST : FIND THE RED DOT !

TEST : FIND THE RED DOT !



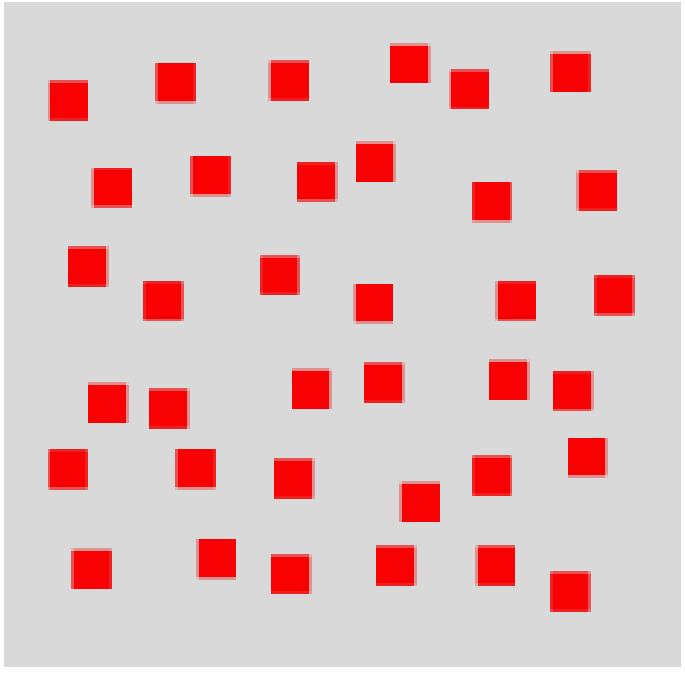
TEST : FIND THE RED DOT !

TEST : FIND THE RED DOT !



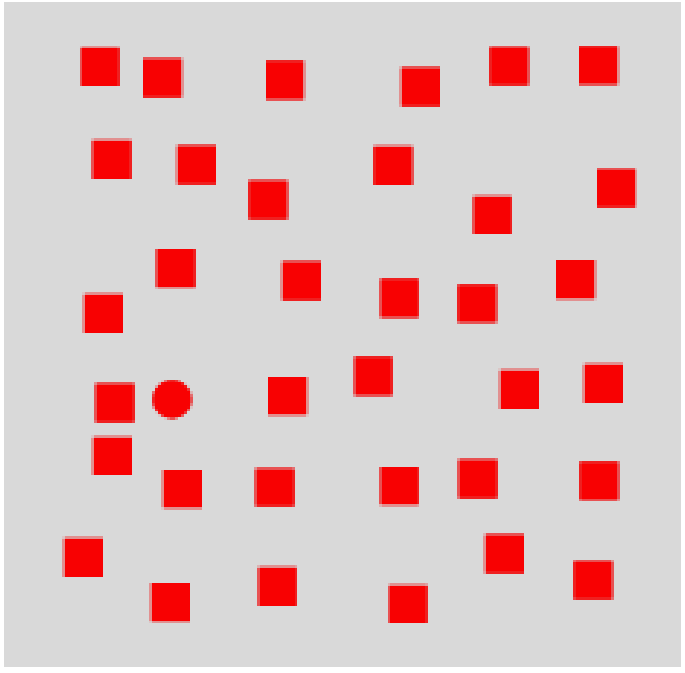
TEST : FIND THE RED DOT !

TEST : FIND THE RED DOT !



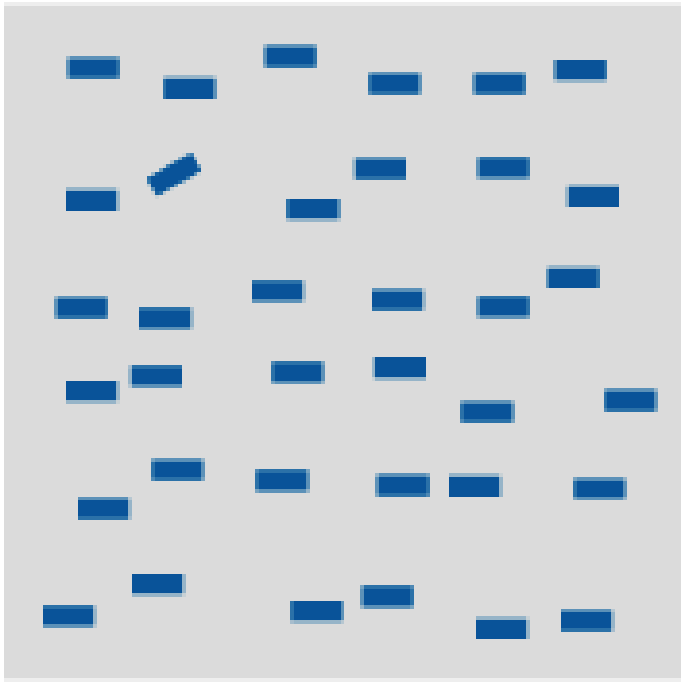
TEST : FIND THE RED DOT !

TEST : FIND THE RED DOT !



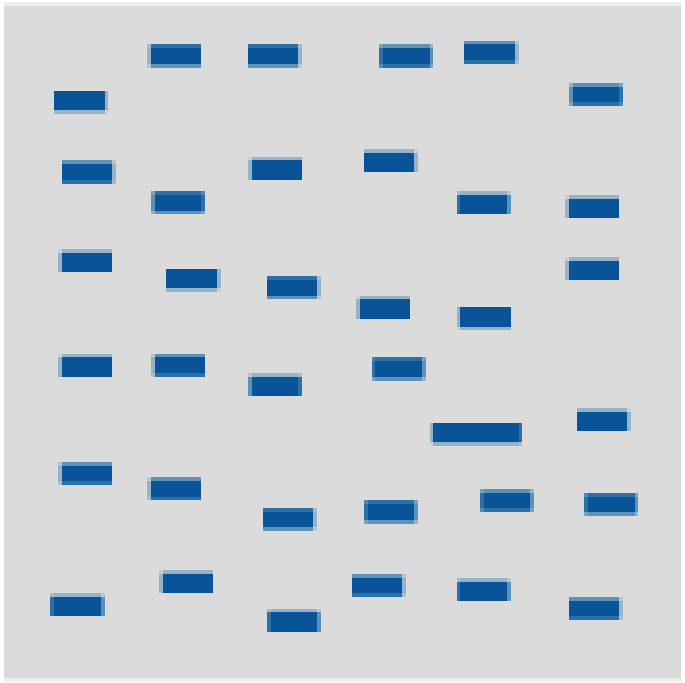
HARDER : IS THERE A "STRANGER" ?

HARDER : IS THERE A "STRANGER" ?



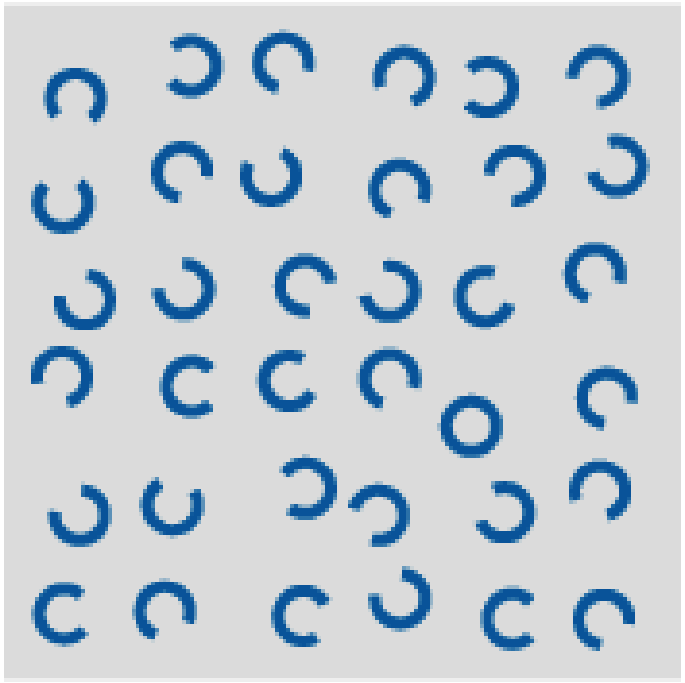
HARDER : IS THERE A "STRANGER" ?

HARDER : IS THERE A "STRANGER" ?



HARDER : IS THERE A "STRANGER" ?

HARDER : IS THERE A "STRANGER" ?



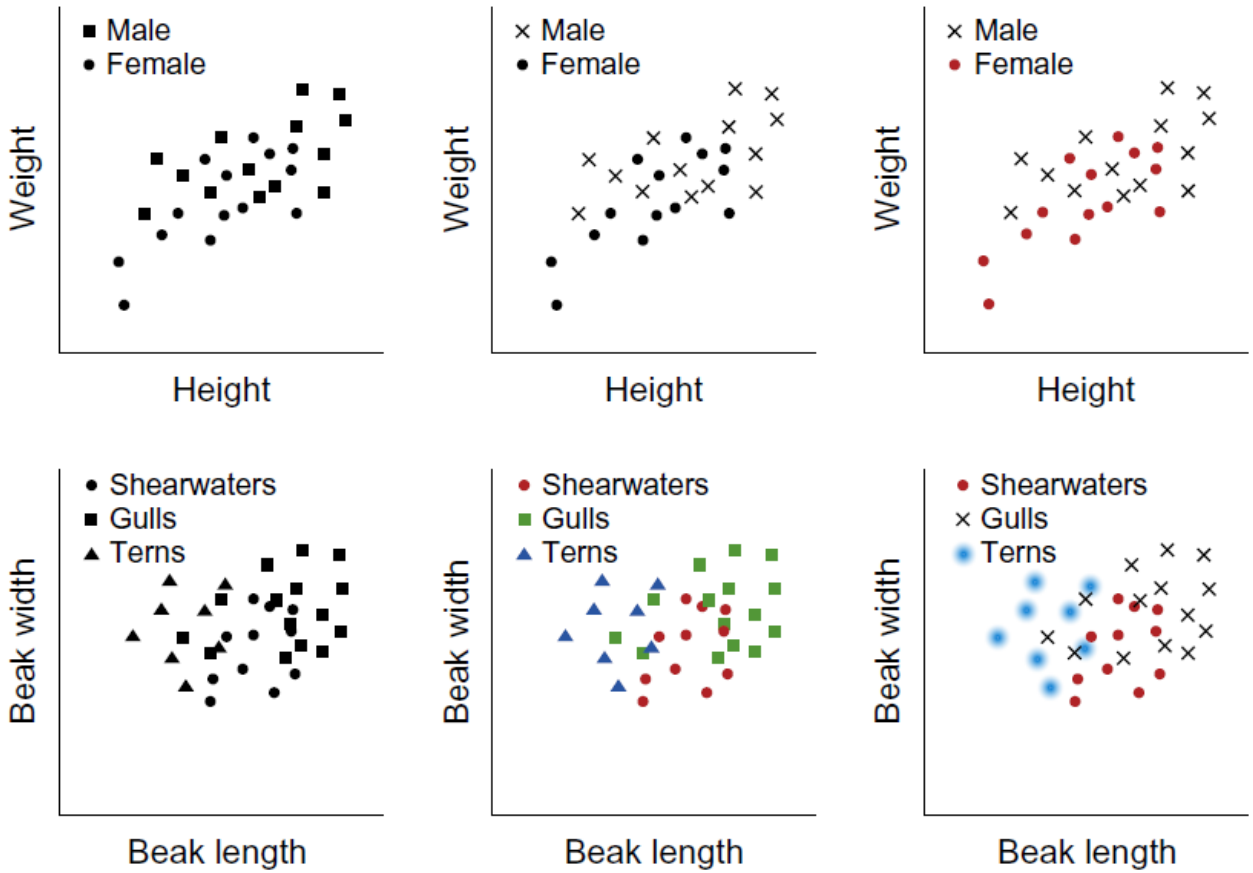
HARDER : IS THERE A "STRANGER" ?

That wasn't easy

- Preattentive concept, Treisman (1985)
- Some visual elements or patterns are detected immediately
- But there may be interferences (colour and form)
- Very useful (detection, explanatory and presentation)

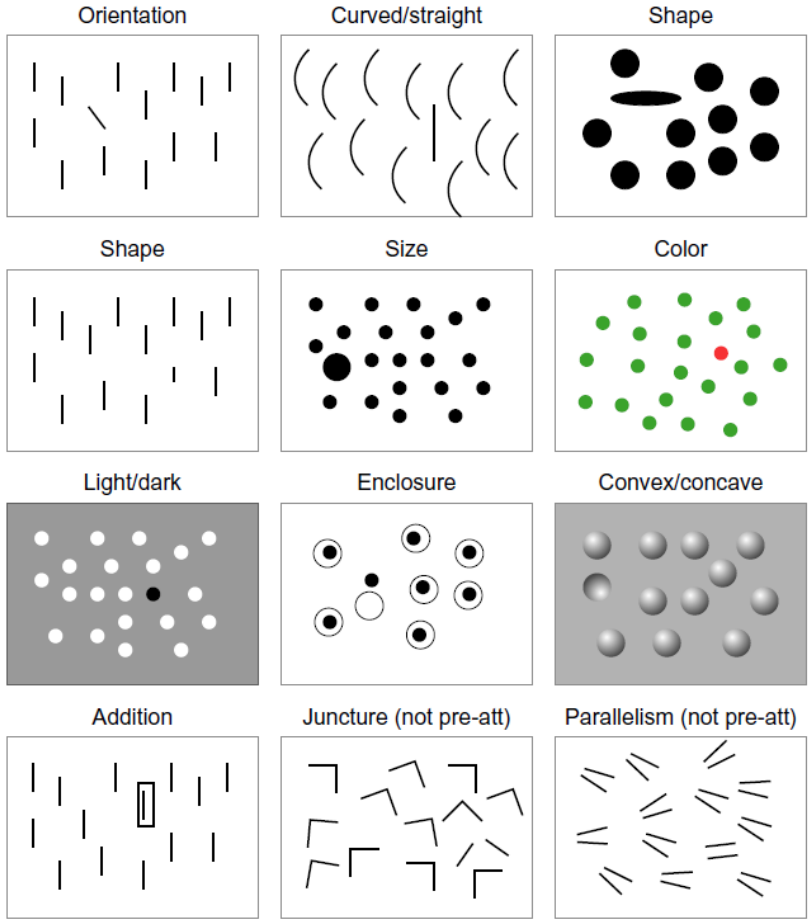
Helpful to highlight a message !

Too much variation doesn't help



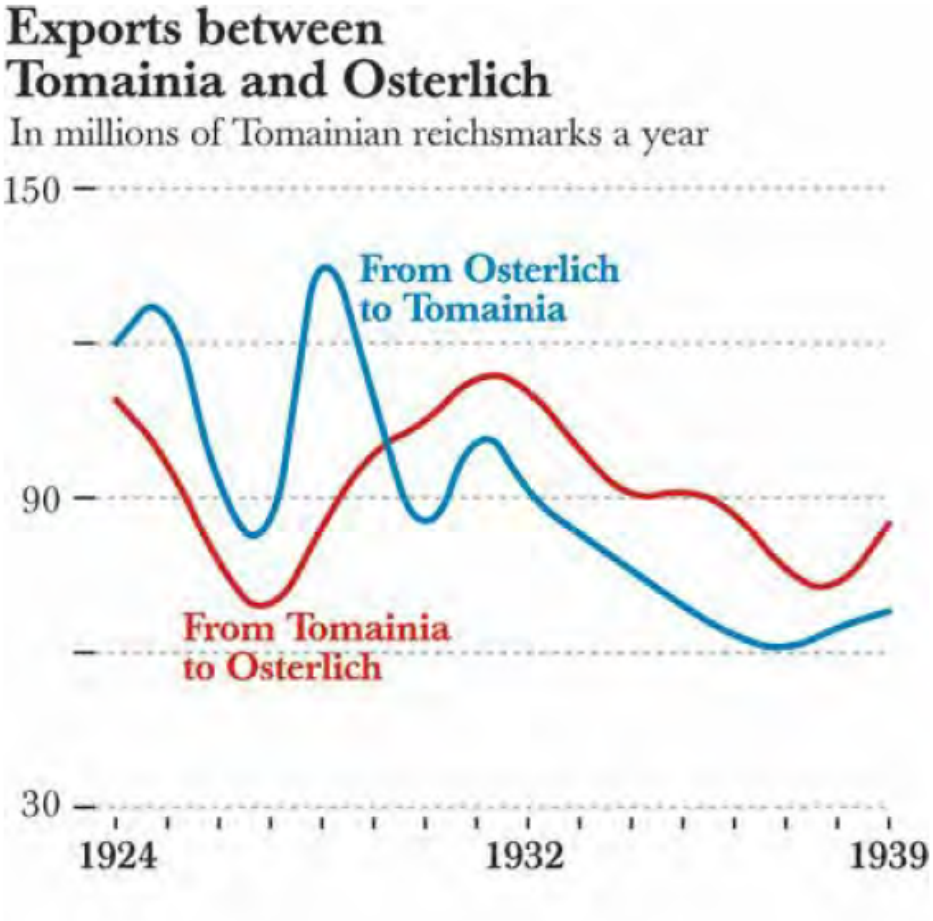
From Ware (2012)

Most preattentive visual variables



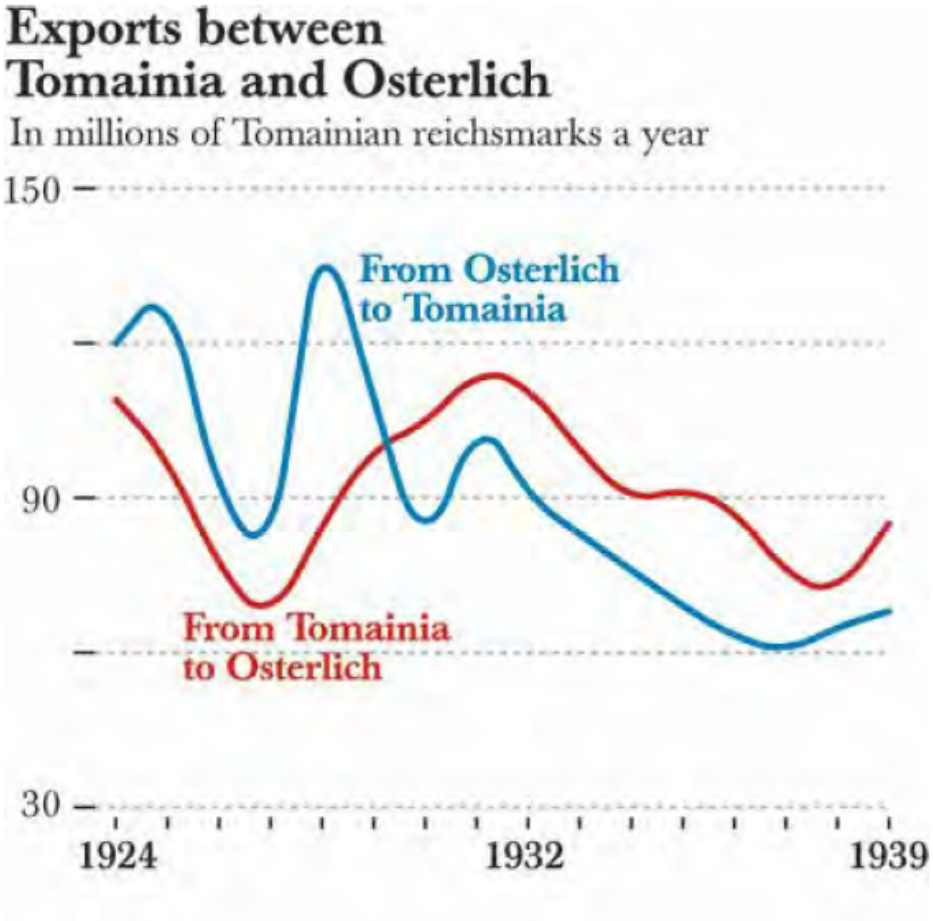
From Ware (2012)

VISUAL PERCEPTION AND LINES



From Cairo (2012)

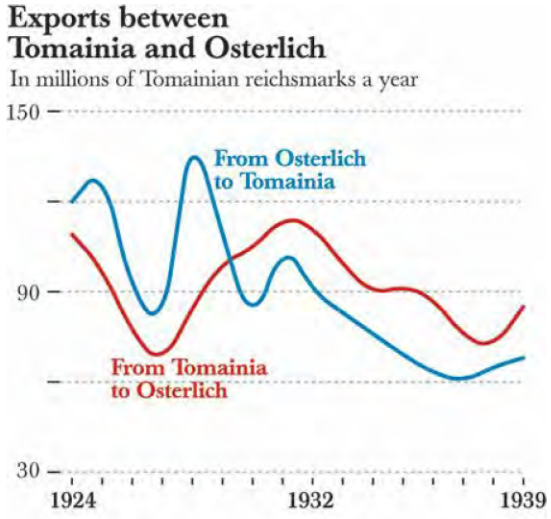
VISUAL PERCEPTION AND LINES



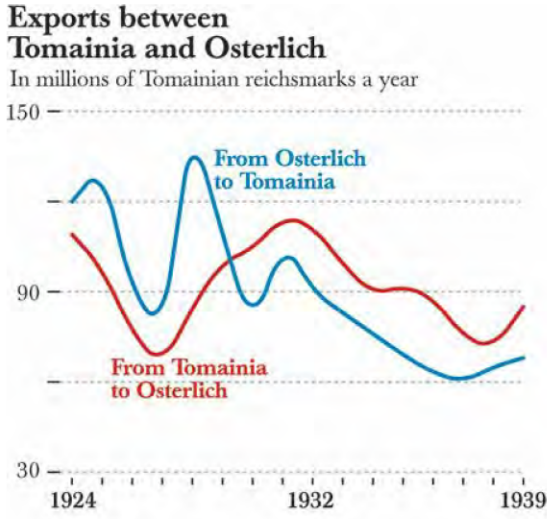
When was the biggest negative (positive) difference ?

From Cairo (2012)

VISUAL PERCEPTION AND LINES

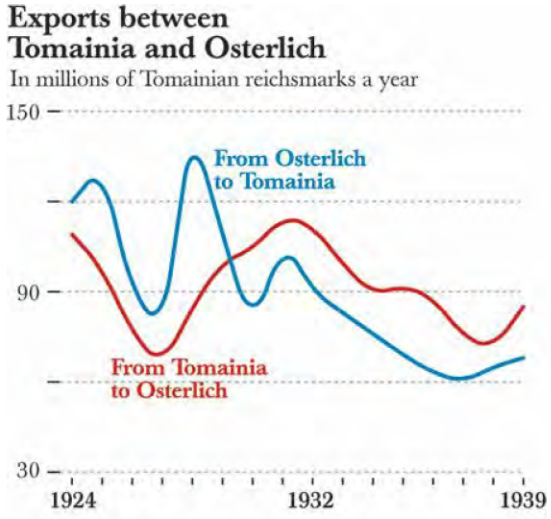


VISUAL PERCEPTION AND LINES

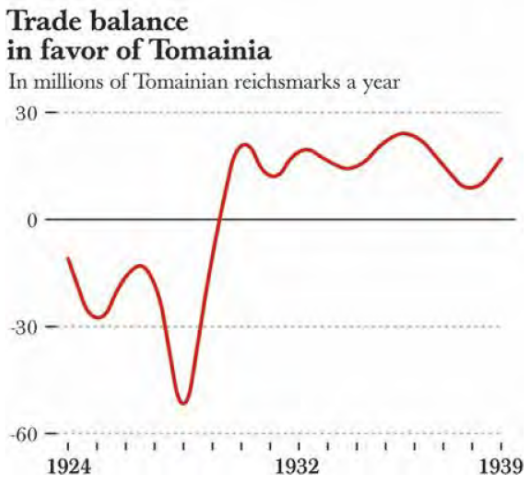


When was the biggest negative (positive) difference ?

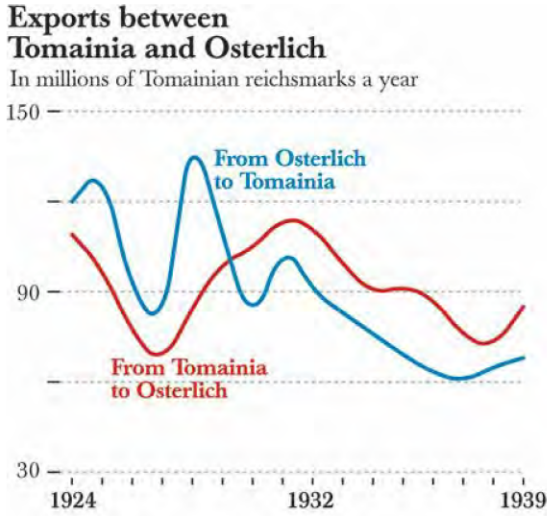
VISUAL PERCEPTION AND LINES



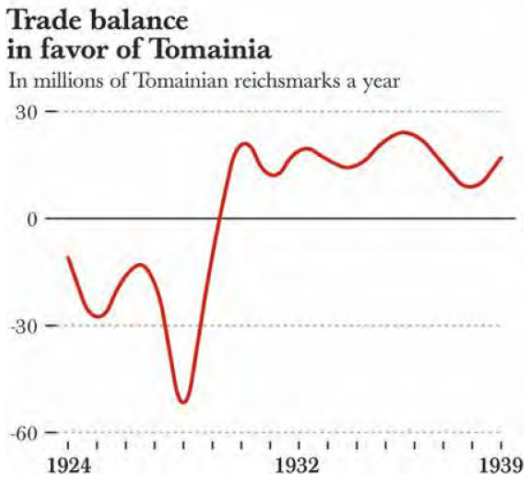
When was the biggest negative (positive) difference ?



VISUAL PERCEPTION AND LINES

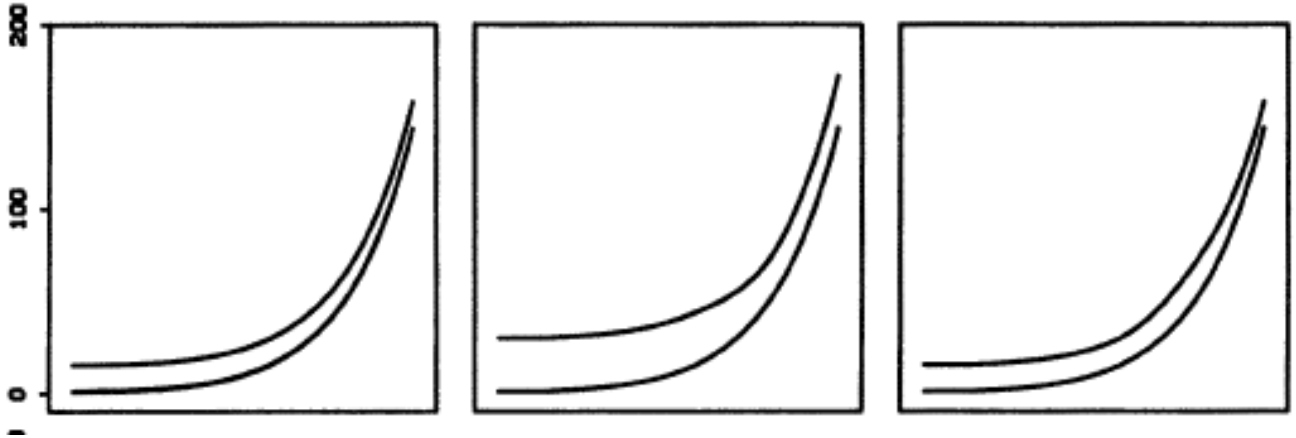


When was the biggest negative (positive) difference ?

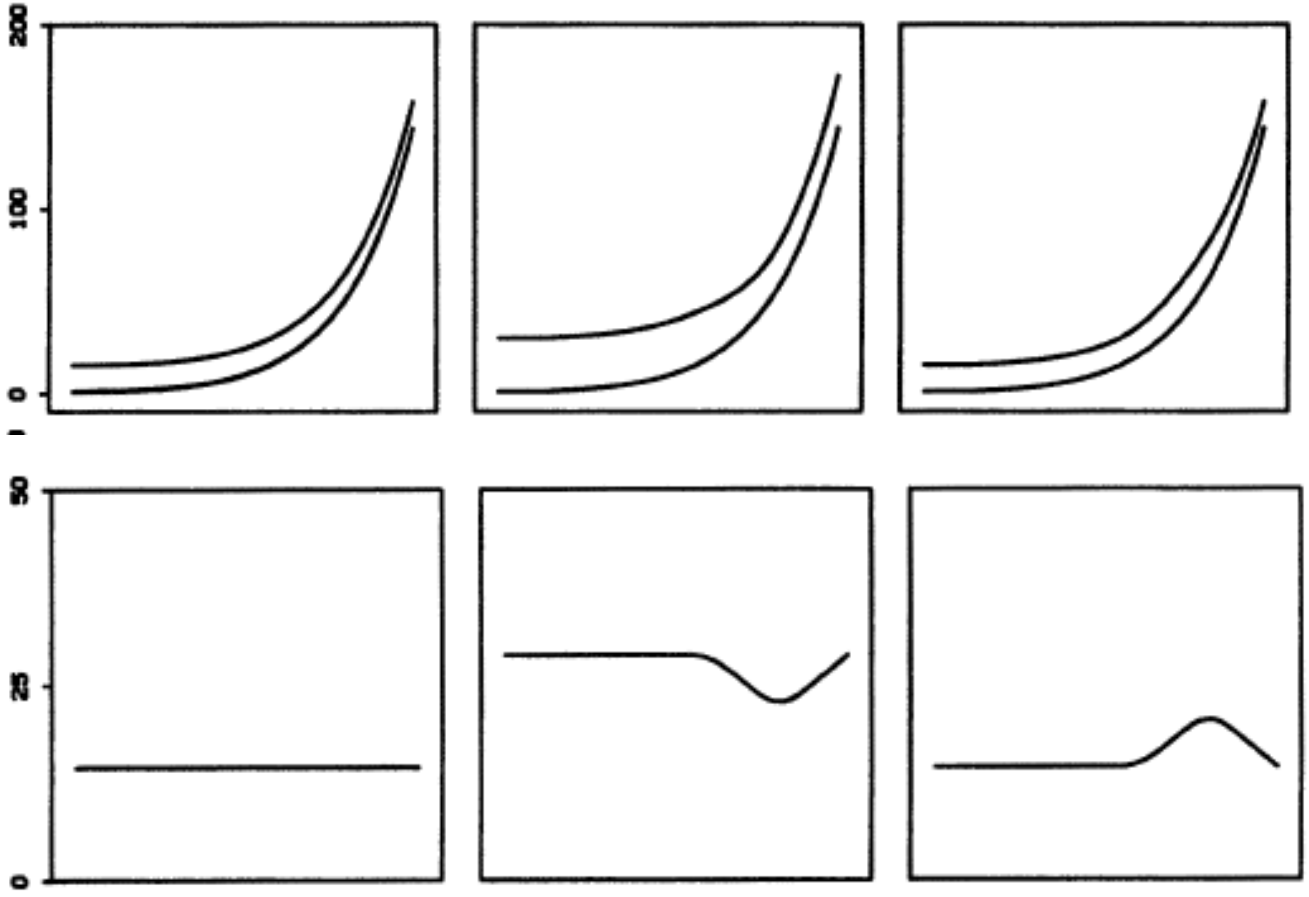


From Cairo (2012)

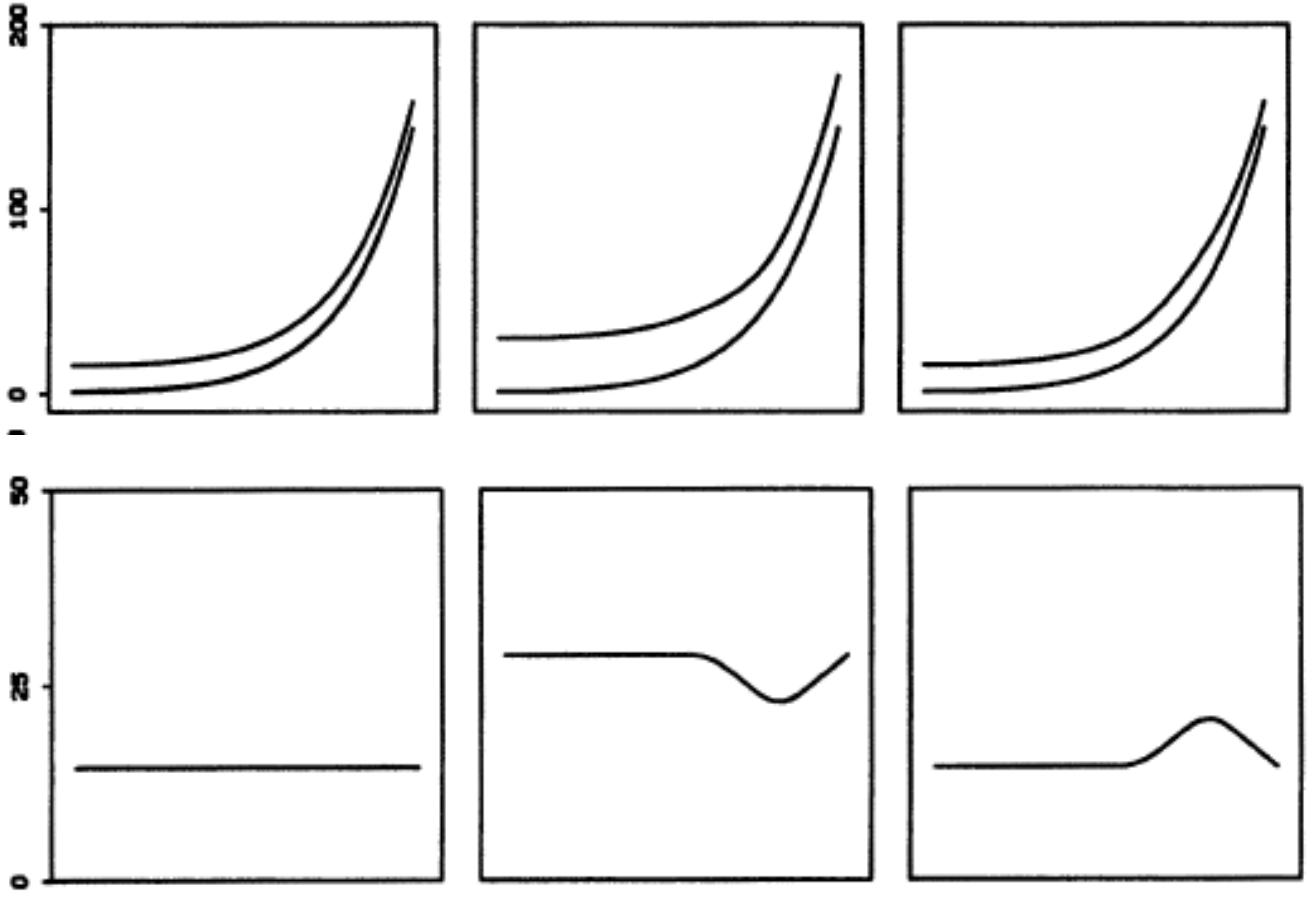
The Cleveland-McGill effect



The Cleveland-McGill effect

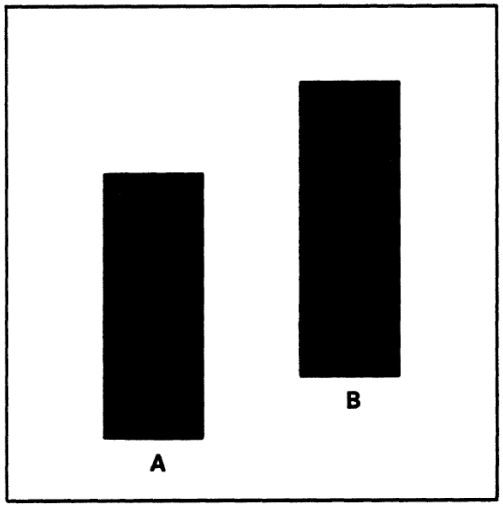


The Cleveland-McGill effect

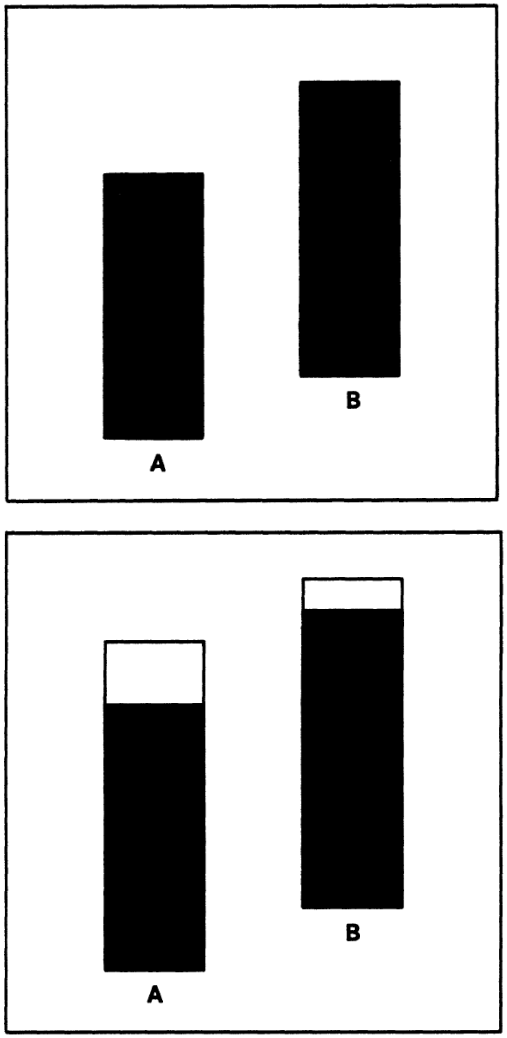


From Cleveland and McGill (1984)

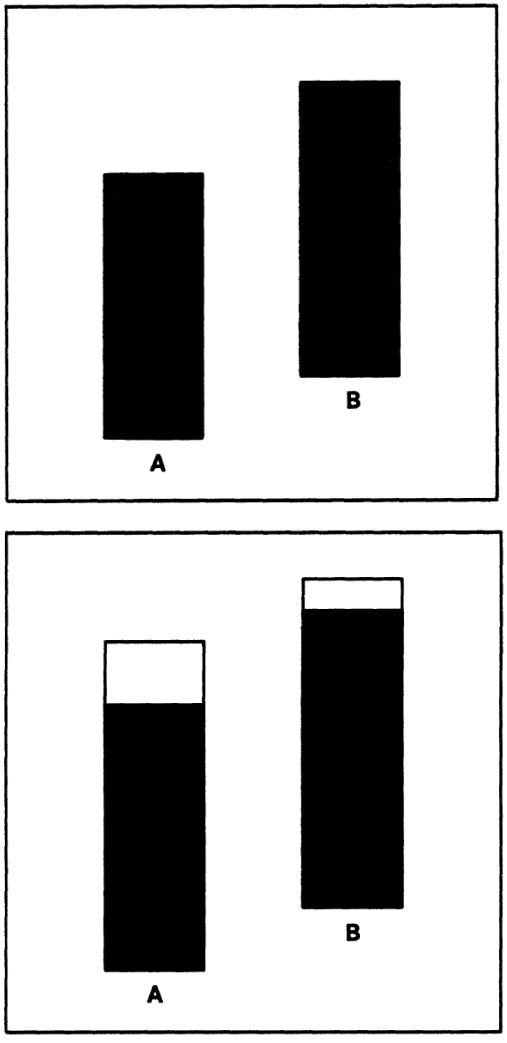
Weber's law and framed boxes



Weber's law and framed boxes

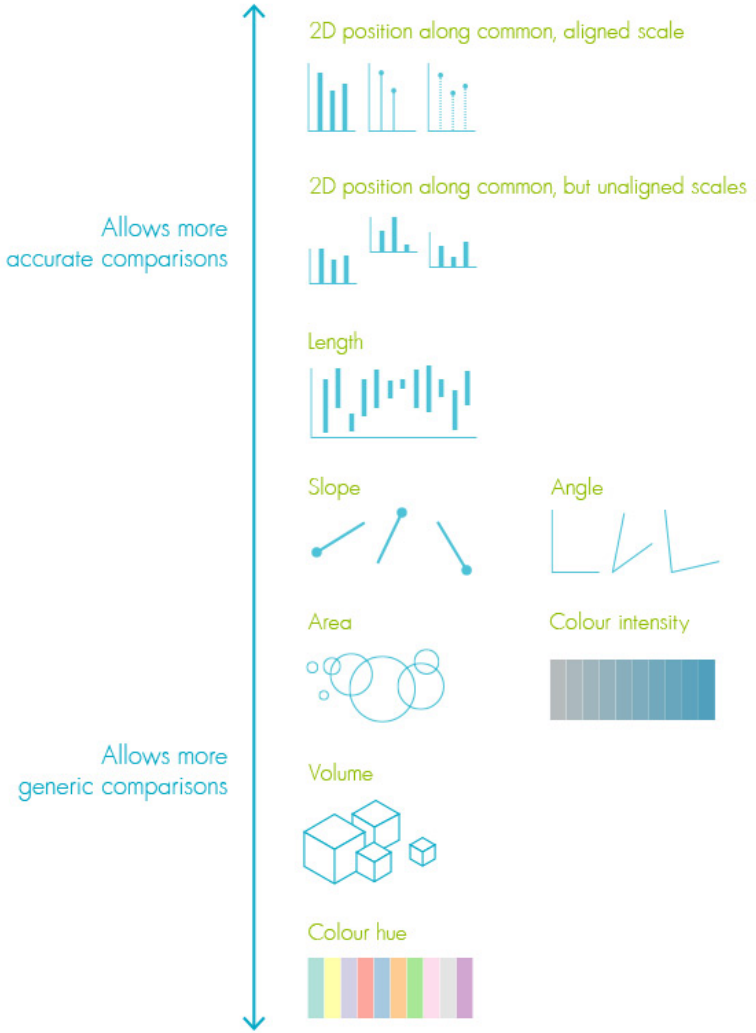


Weber's law and framed boxes



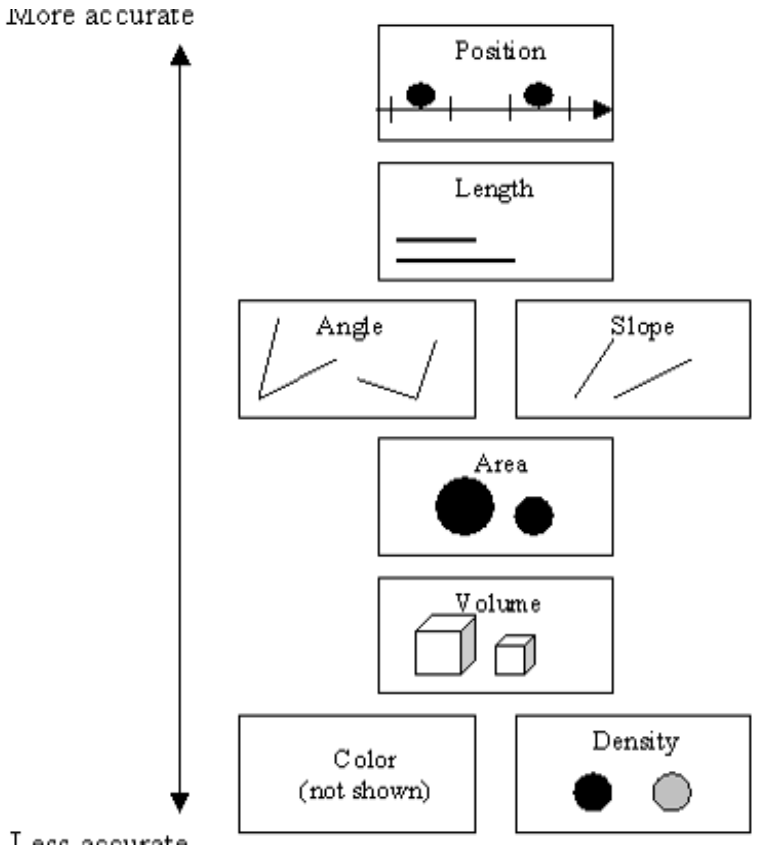
From Cleveland and McGill (1984)

The Cleveland-McGill Scale



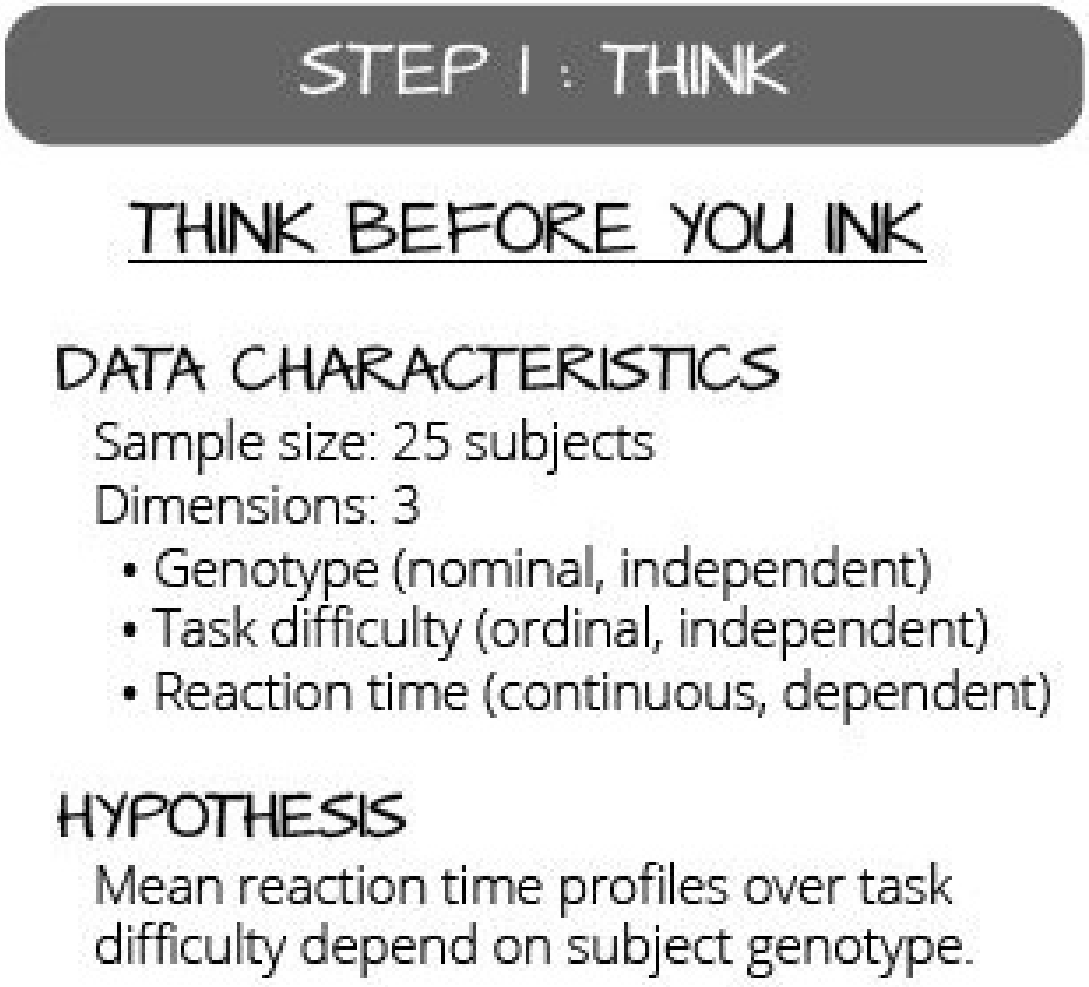
Cleveland and McGill (1984)

THE CLEVELAND-MCGILL SCALE



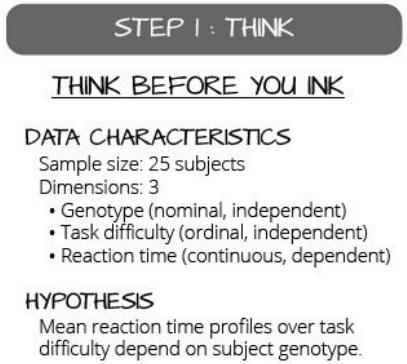
Source `http://hci12.cs.umd.edu/trs/99-20/99-20.html`

THE DATAVIZ PROCESS



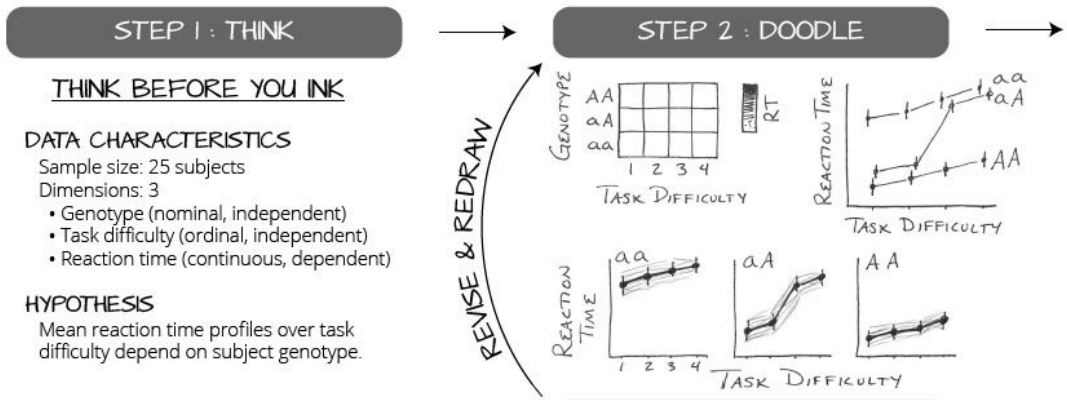
From Allen and Erhardt (2016a)

THE DATAVIZ PROCESS



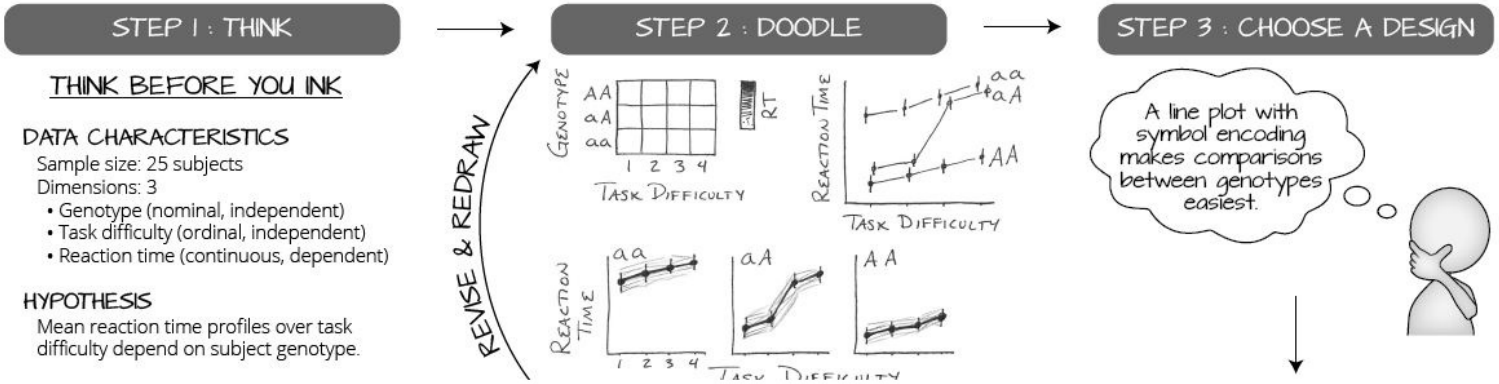
From Allen and Erhardt (2016a)

THE DATAVIZ PROCESS



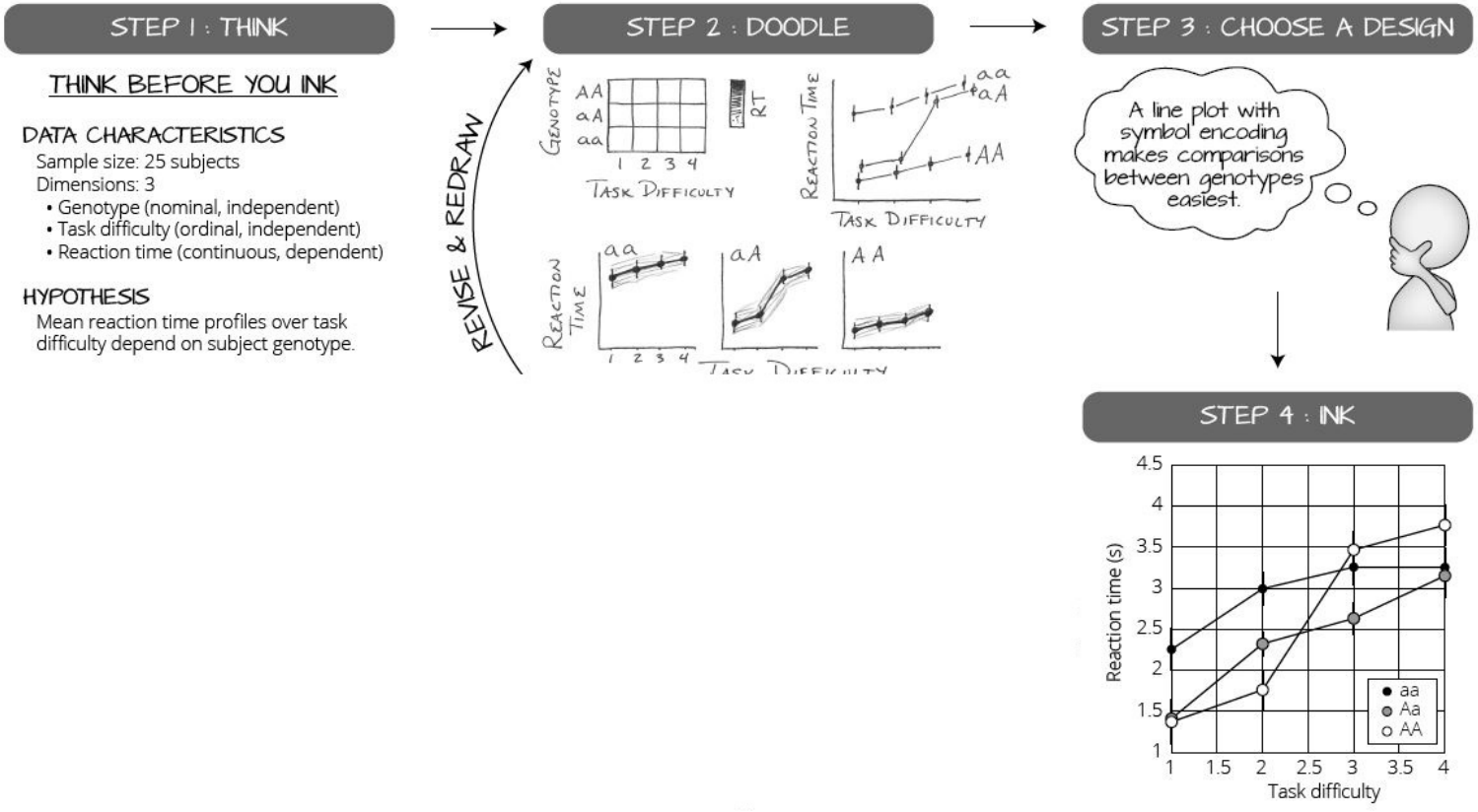
From Allen and Erhardt (2016a)

THE DATAVIZ PROCESS



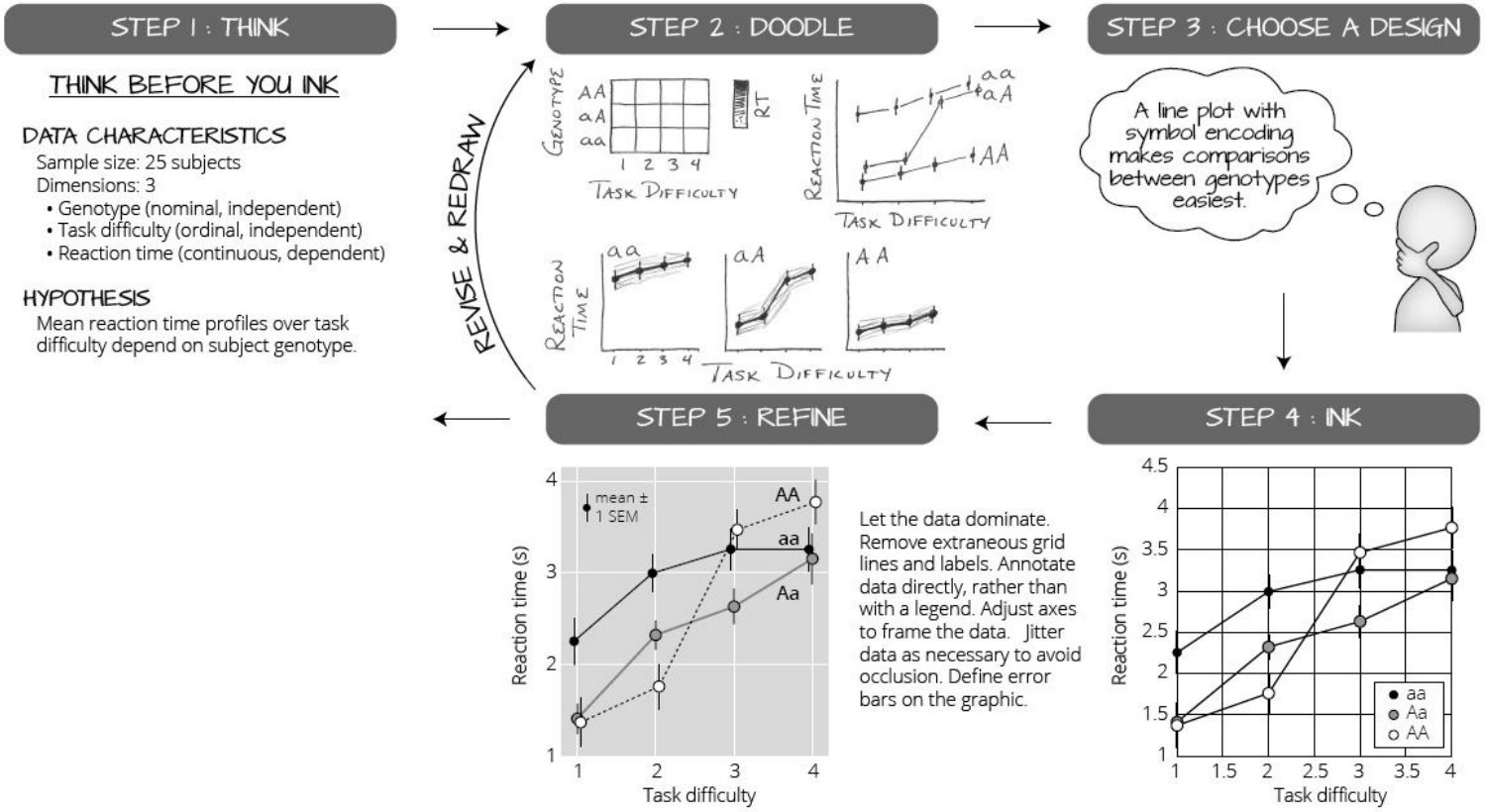
From Allen and Erhardt (2016a)

THE DATAVIZ PROCESS



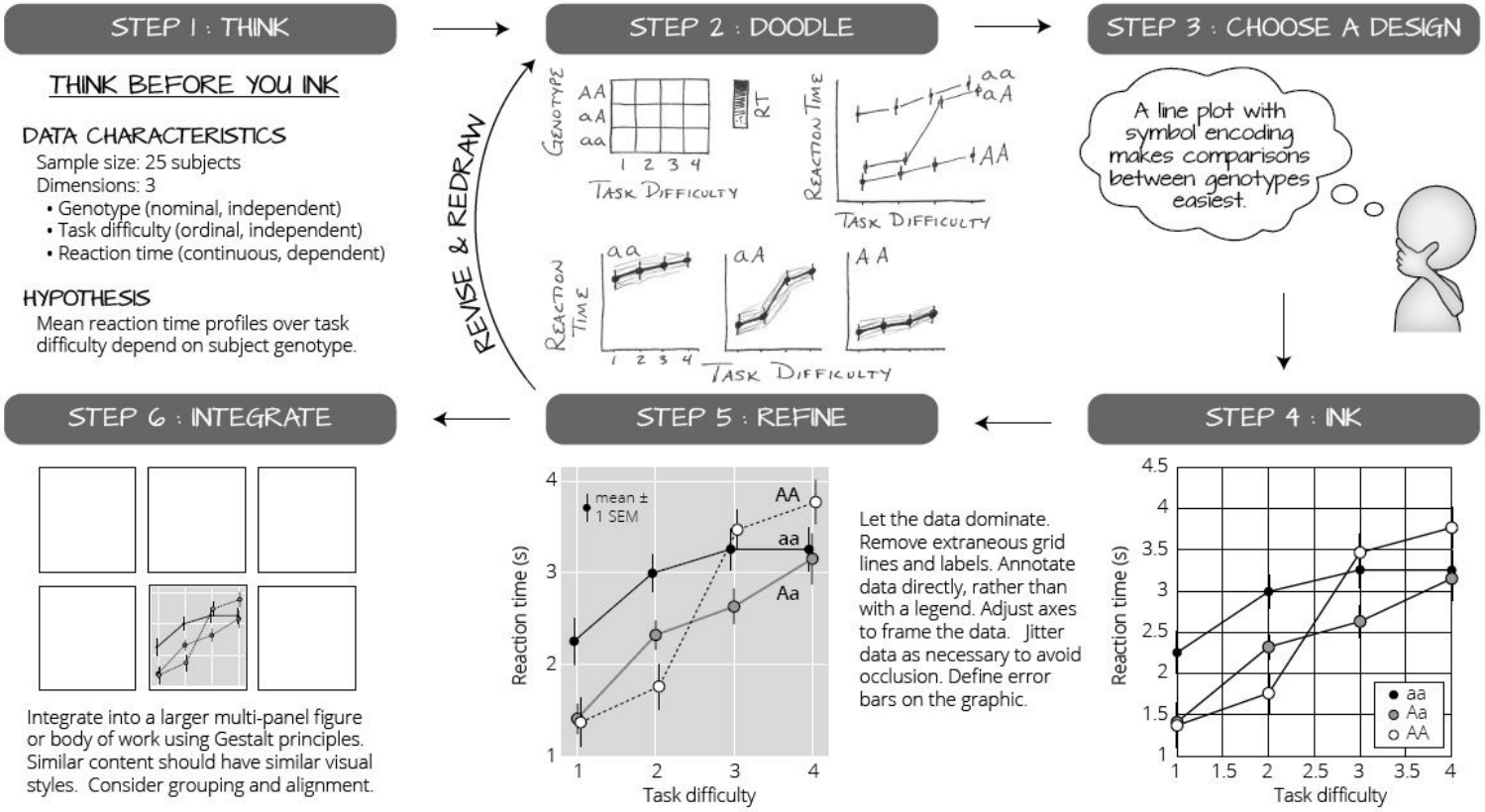
From Allen and Erhardt (2016a)

THE DATAVIZ PROCESS



From Allen and Erhardt (2016a)

THE DATAVIZ PROCESS



From Allen and Erhardt (2016a)

Partial conclusion

Gordon and Finch (2014) gives some nice principles

Partial conclusion

Gordon and Finch (2014) gives some nice principles

1. Show the data clearly

Partial conclusion

Gordon and Finch (2014) gives some nice principles

1. Show the data clearly
2. Use simplicity in design

Partial conclusion

Gordon and Finch (2014) gives some nice principles

1. Show the data clearly
2. Use simplicity in design
3. Use good alignment on a common scale for quantities to be compared

Partial conclusion

Gordon and Finch (2014) gives some nice principles

1. Show the data clearly
2. Use simplicity in design
3. Use good alignment on a common scale for quantities to be compared
4. Keep visual encoding transparent

Partial conclusion

Gordon and Finch (2014) gives some nice principles

1. Show the data clearly
2. Use simplicity in design
3. Use good alignment on a common scale for quantities to be compared
4. Keep visual encoding transparent
5. Use graphical forms consistent with those principles

Partial conclusion

Gordon and Finch (2014) gives some nice principles

1. Show the data clearly
2. Use simplicity in design
3. Use good alignment on a common scale for quantities to be compared
4. Keep visual encoding transparent
5. Use graphical forms consistent with those principles

We may add some others (use preattentive elements, integrity, ...)

CHECKLIST

Gordon and Finch (2014) provides a checklist

A checklist for good graphical practice

How clear is your purpose in communication?

- ☐ What relationships or patterns can you identify in the graph?
- ☐ Are these the relationships or patterns you intended to represent?
- ☐ Can the viewer identify the patterns you wish to illustrate?
- ☐ Are the important comparisons you wish to show salient?

Make clarity a high priority.

- ☐ Does the graph have a clear title?
- ☐ Are the axes labelled?
- ☐ Are the units of the variables measured defined?
- ☐ Are the units of observation clear?
- ☐ Is the graph large enough?
- ☐ Would ordering groups or variables plotted improve the graph?
- ☐ Are all the graph labels horizontal?
- ☐ Use points to plot estimates (e.g. means, proportions) rather than bars.

Choose standard forms fit for your purpose.

- ☐ Use bars around points to indicate the precision of the estimates.
- ☐ Plot the estimates of interest (e.g. mean differences with confidence intervals) rather than standard summary statistics (group means).
- ☐ Plot inferences to support stories about models.
- ☐ Plot data to support stories about distributions and variation.

Consider detection issues.

- ☐ Can all the data points be seen?
- ☐ Are patterns in the data clear?

Checklist

Gordon and Finch (2014) provides a checklist

- ☐ Are the fonts large enough?
- ☐ Would it help to use jittering, or another form of representing multiple, identical values?

Would panels help?

- ☐ Are there grouping variables that can be used to panel the graph?
- ☐ Do the grouping variables correspond to the variation of interest?
- ☐ Would additional panels help?

Align quantities to be compared on a common scale.

- ☐ Has distortion of the data been avoided by using the same scales for the same measurement?
- ☐ Are measurements made on the same scale plotted on the same scale?
- ☐ Would transposition improve the graph?

Does the graph have grid lines?

- ☐ Light grey grid lines will help with accurate interpretation.

Are all the elements of the graph defined?

- ☐ What do points on the graph correspond to?
- ☐ Are estimates (e.g. means, proportions) plotted on the graph clearly defined?
- ☐ Are bars around points on the graph clearly defined?

How much decoding work does the viewer have to do?

- ☐ Is it easy for someone unfamiliar with your data to interpret your graph?
- ☐ Does the graph stand alone?
- ☐ Try it on a friend!

References I

Allen, E. A. and Erhardt, E. B. (2016a). *Handbook of Psychophysiology,*. Cambridge University Press, 4th edition edition.

Allen, E. A. and Erhardt, E. B. (2016b). Visualizing Scientific Data. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of Psychophysiology*, pages 679–697. Cambridge University Press, 4 edition.

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1) :17–21.

Bertin, J. (1970). La graphique. *Communications*, 15(1).

Bertin, J. (1981). Théorie matricielle de la graphique. *Communication et langages*, 48(1) :62–74.

Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.-K., Swayne, D. F., and Wickham, H. (2009). Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, 367(1906) :4361–4383.

Cairo, A. (2012). *The Functional Art : An introduction to information graphics and visualization*. Voices That Matter. Pearson Education.

References II

Chen, C.-h., Härdle, W. K., and Unwin, A. (2007). *Handbook of data visualization*. Springer Science & Business Media.

Cleveland, W. S. (1994). *The Elements of Graphing Data*. Hobart Press, Summit : NJ, 2 edition.

Cleveland, W. S. and McGill, R. (1984). Graphical perception : Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387) :531–554.

Dix, A. and Ellis, G. (1998). Starting simple - adding value to static visualisation through simple interaction. In Eds. T. Catarci, M. F. Costabile, G. S. and Tarantino, L., editors, *Proceedings of Advanced Visual Interfaces*, pages 124–134. L'Aquila, Italy, ACM Press.

Few, S. (2008). Practical rules for using color in charts. *Visual Business Intelligence Newsletter*, (11).

Gelman, A. (2004). Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics*, 13(4).

Gordon, I. and Finch, S. (2014). Statistician heal thyself : Have we lost the plot ? *Journal of Computational and Graphical Statistics*, 24(4) :1210–1229.

References III

Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting things in order : an introduction to the r package seriation. *Journal of Statistical Software*, 25(3) :1–34.

Matejka, J. and Fitzmaurice, G. (2017). Same stats, different graphs : generating datasets with varied appearance and identical statistics through simulated annealing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1290–1294. ACM.

Munzner, T. (2014). *Visualization Analysis and Design*. AK Peters Visualization Series. A K Peters/CRC Press, 1 edition.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31(2) :156–177.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press, 2 edition.

Ware, C. (2012). *Information visualization : perception for design*. Elsevier.

Wong, D. M. (2010). *The Wall Street Journal guide to information graphics : The dos and don'ts of presenting data, facts, and figures*. WW Norton.