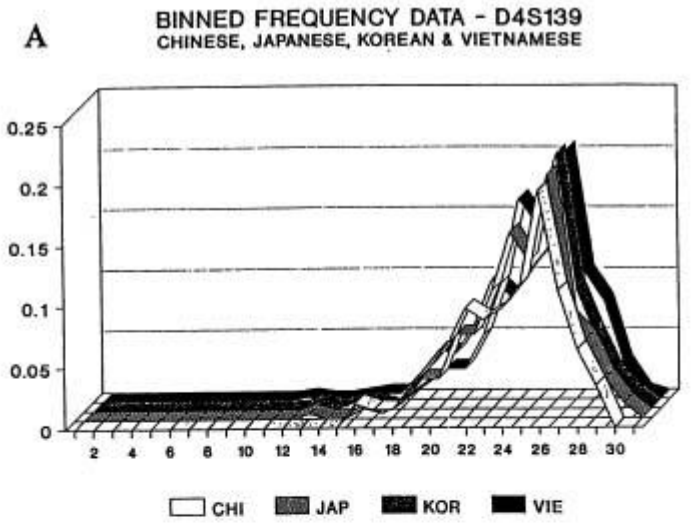
The top ten worst graphs

1. Roeder K (1994) DNA fingerprinting: A review of the controversy (with discussion). Statistical Science 9:222-278, Figure 4
[The article | The figure | Discussion]



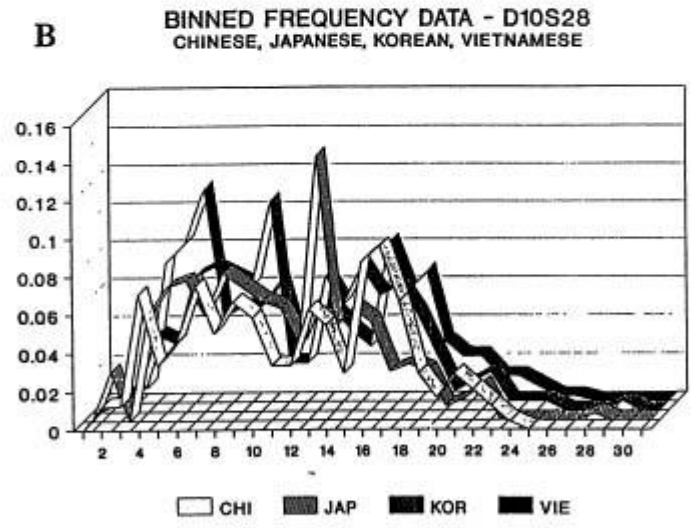


FIG. 4. *Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the* 30 *bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean and Vietnamese are* 103, 125, 93 *and* 215 *for* D4S139 *and* 120, 137, 100 *and* 193 *for* D10S28. *The horizontal axis is the bin number; bins are not of equal length.*

What's wrong with this one?

Curves rendered as ribbons? The 3-dimensional rendering of the curves is entirely gratuitous.

What should have been done?

Without a doubt, it's difficult to display multiple curves simultaneously and ensure that the individual curves may be seen. Colors would be nice, but if color is not allowed, four different line types (solid, dashed, dotted, dash-dotted) might work.

2. Wittke-Thompson JK, Pluzhnikov A, Cox NJ (2005) Rational inferences about departures from Hardy-Weinberg equilibrium. American Journal of Human Genetics 76:967-986, Figure 1
[The article | Fig 1AB | Fig 1CD | Discussion]

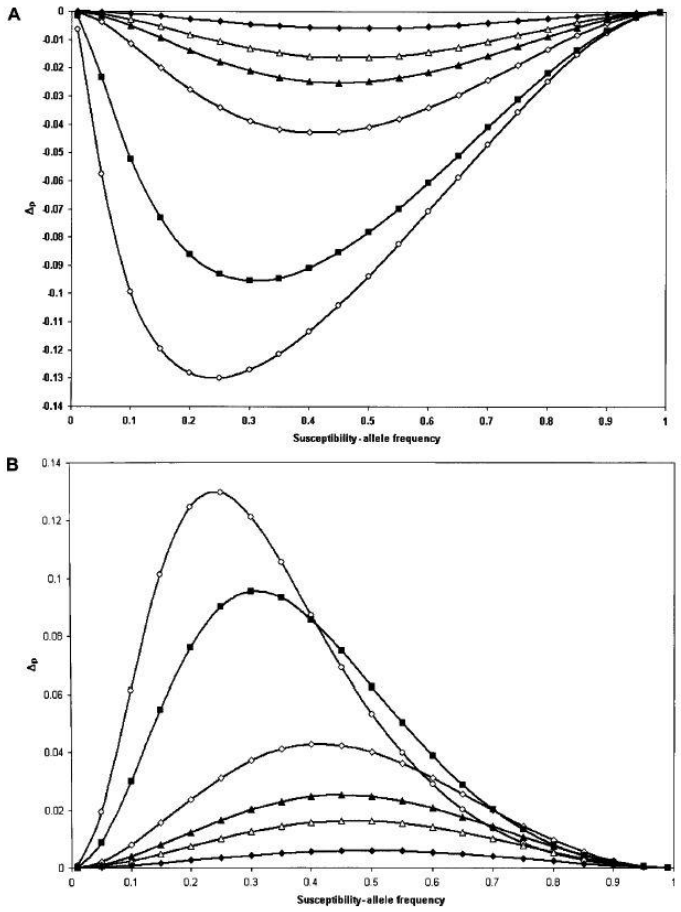
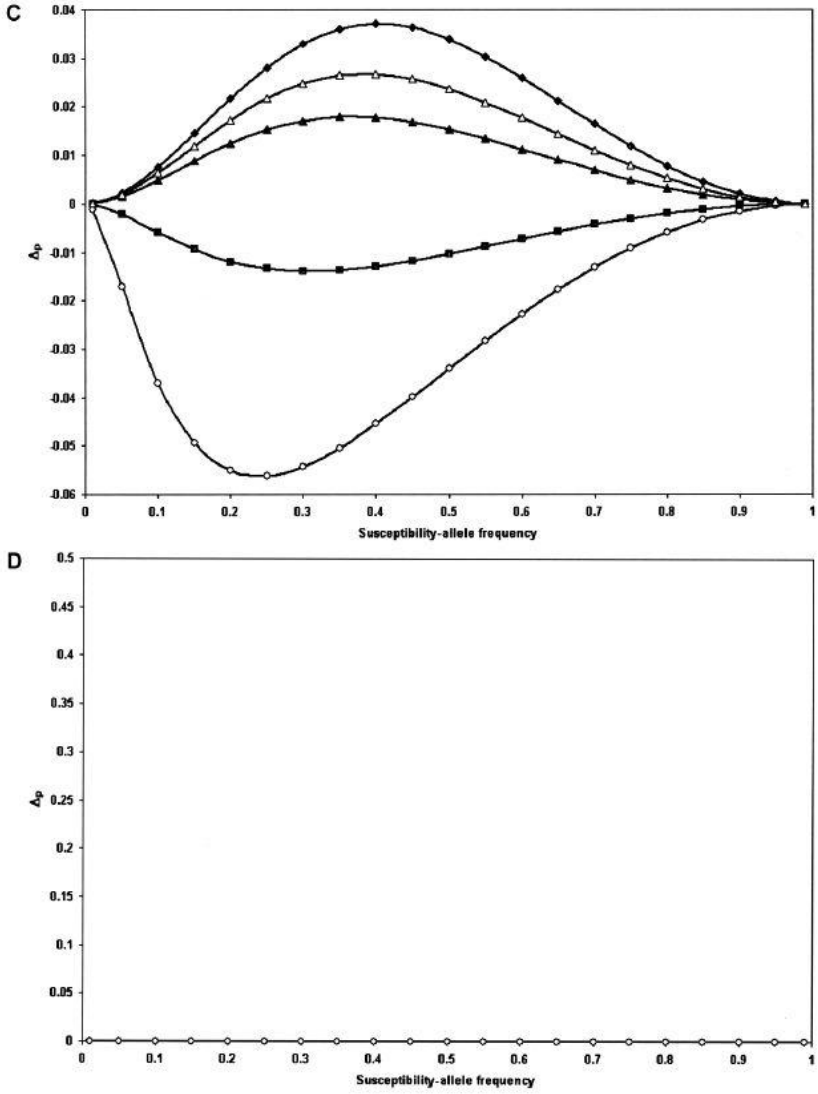


Figure 1 Δ_γ plotted versus the susceptibility-allele frequency for patients. *A*, *B*, and *D*, Data points are as follows: $\gamma = 1.1$ (*blackened diamonds*), $\gamma = 1.3$ (*unblackened triangles*), $\gamma = 1.5$ (*blackened triangles*), $\gamma = 2$ (*unblackened diamonds*), $\gamma = 5$ (*blackened squares*), and $\gamma = 10$ (*unblackened circles*). *A*, Dominant model. *B*, Recessive model. *C*, Additive model. Since $\gamma < 2$ would not satisfy our definition of an additive model as $\gamma = 2\beta$ and $\beta > 1$, the data points in *C* are as follows: $\gamma = 2.2$ ($\beta = 1.1$) (*blackened diamonds*), $\gamma = 2.6$ ($\beta = 1.3$) (*unblackened triangles*), $\gamma = 3$ ($\beta = 1.5$) (*blackened triangles*), $\gamma = 5$ (*blackened squares*), $\gamma = 2$ (*unblackened diamonds*). *D*, Multiplicative model.



What's wrong with this one?

If you want your graph to be published in large format, it seems that the *American Journal of Human Genetics* is the place to which you should aim. This figure spans two pages. Panel D is most interesting; it takes a while to identify that there is any information there at all.

What should have been done?

Figure 1D could have been discarded completely, and the whole figure shouldn't take up more than half a page.

3. Epstein MP, Satten GA (2003) Inference on haplotype effects in case-control studies using unphased genotype data. American Journal of Human Genetics 73:1316-1329, Figure 1
[The article | The figure | Discussion]

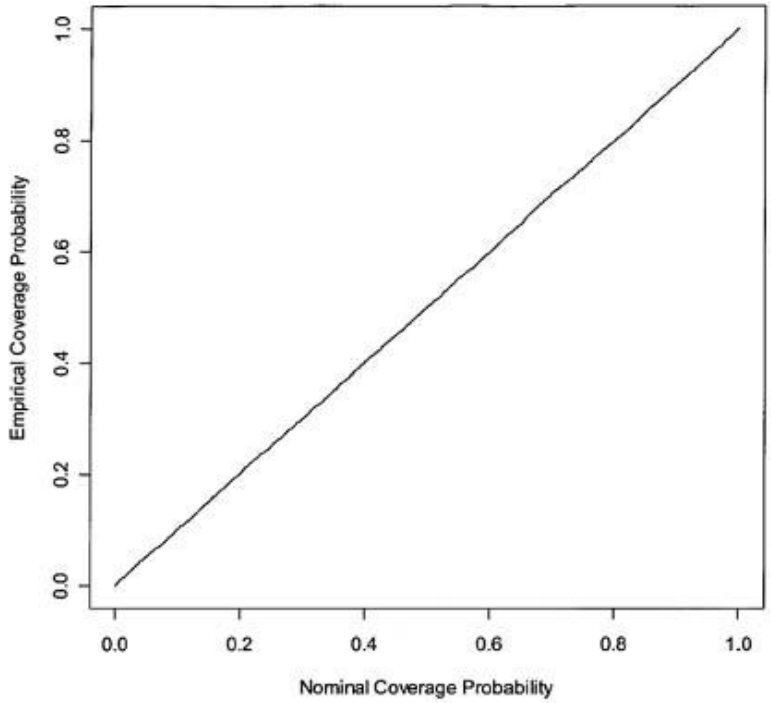


Figure 1 Empirical coverage of CIs for the relative-risk parameter β of haplotype 01100. Results are based on 10,000 simulated data sets with the same haplotype frequencies as the FUSION data. Haplotype 01100 has a multiplicative effect on disease risk, with $\beta = 0.35$.

What's wrong with this one?

This half-page plot contains little information, and all of the interesting action is way in the lower-left corner.

What should have been done?

Plot either the percent difference between empirical and nominal, and just for the region of interest, or just say that the empirical results exactly matched the nominal ones.

4. Mykland P, Tierney L, Yu B (1995) Regeneration in Markov chain samplers. Journal of the American Statistical Association 90:233-241, Figure 1
[The article | The figure | Discussion]

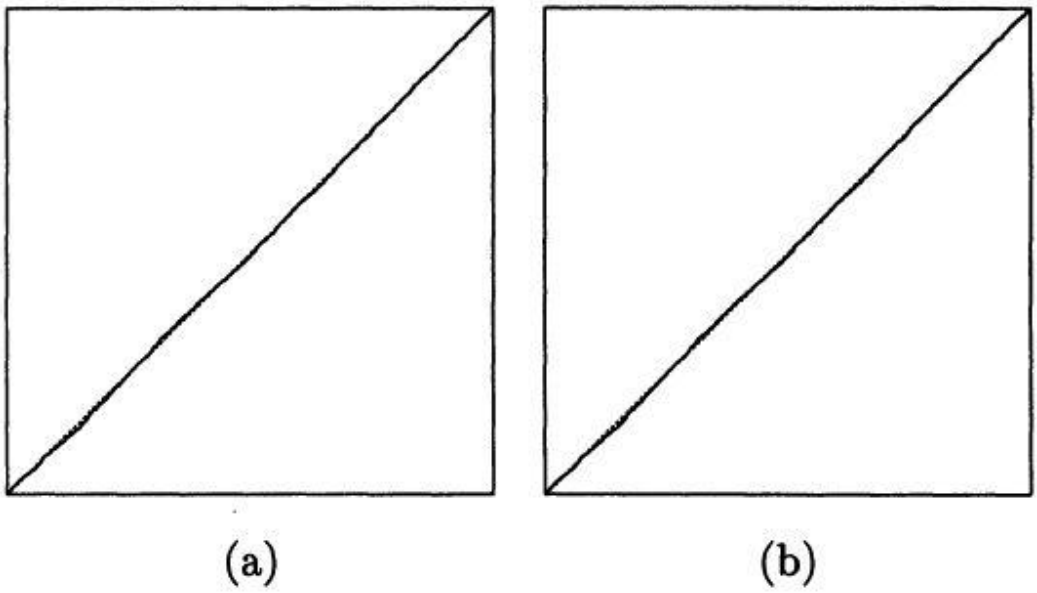


Figure 1. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Pump Failure Data Based on Runs of Length 5,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

What's wrong with this one?

This is like the last one: it contains almost no information. It's redundant, but so funny we couldn't help but include it.

What should have been done?

Plot percent differences, or just say the results are indistinguishable.

5. Hummer BT, Li XL, Hassel BA (2001) Role for p53 in gene induction by double-stranded RNA. J Virol 75:7774-7777, Figure 4
[The article | The figure | Discussion]

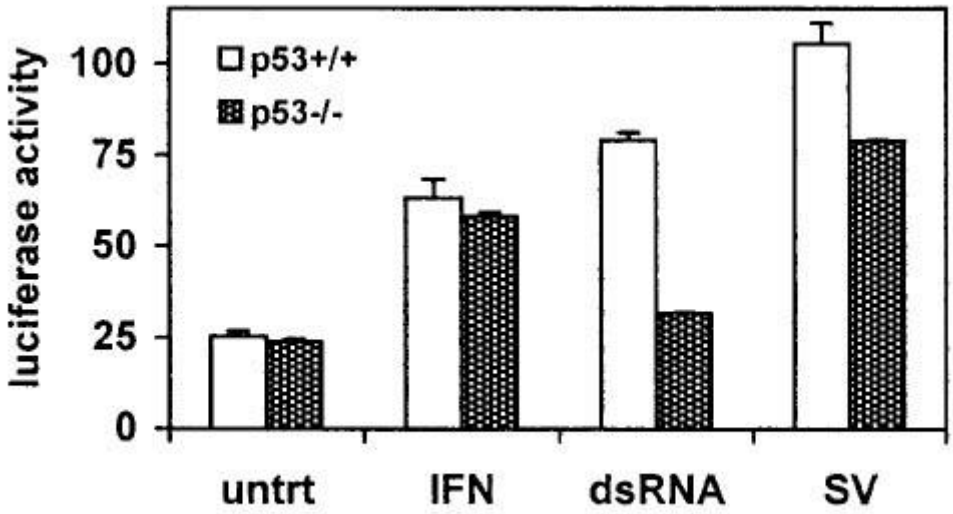


FIG. 4. ISG15 promoter activity mimics endogenous ISG15 mRNA regulation by p53, dsRNA, and virus. Cells (6×10^5 HCT 116) were seeded in 32-mm plates and allowed to attach overnight. Cells were transfected with 500 ng of pGL3/ISG15-Luc, 50 ng of pRL null (Promega), and 450 ng of pcDNA3 for carrier DNA by using Lipofectamine Plus (Life Technologies) following the manufacturer's instructions. Twenty-four hours posttransfection, the medium was aspirated and replaced with medium containing either 1,000 U of IFN-α/ml, 50 μg of dsRNA/ml, or Sendai virus (multiplicity of infection, 10). Cells were incubated for 12 h and then lysed, and luciferase assays were performed. Luciferase activity was assessed on 20 μl of each lysate as directed by the supplier (Dual Luciferase Kit, Promega) using a TD 20/20 luminometer (Turner Designs). Luciferase activity is presented as the ratio of firefly activity to renilla activity to control for differences in transfection efficiency. Each data point is the mean of triplicate samples ± the standard error; the data presented are representative of four independent experiments.

What's wrong with this one?

The bars and little antennae represent just three data points each.

What should have been done?

With just three data points in each group, why not just show the data as dots? You could also include line segments at the averages and even confidence intervals...all this in the same amount of space and with less ink.

6. Cawley S, et al. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116:499-509, Figure 1
[The article | The figure | Discussion]

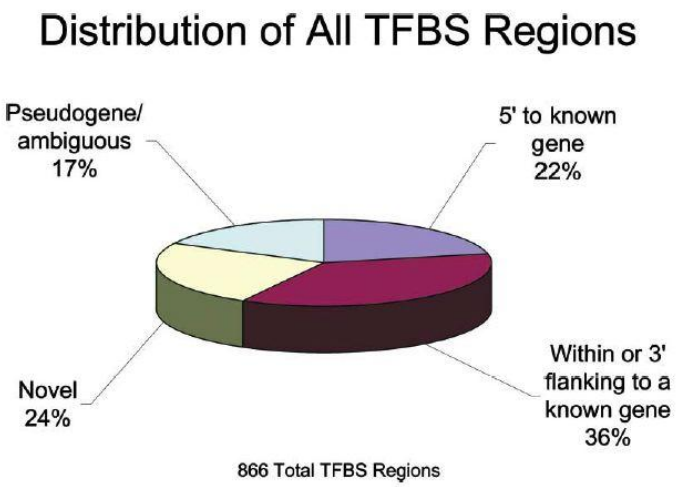


Figure 1. Classification of TFBS Regions
TFBS regions for Sp1, cMyc, and p53 were classified based upon proximity to annotations (RefSeq, Sanger hand-curated annotations, GenBank full-length mRNAs, and Ensembl predicted genes). The proximity was calculated from the center of each TFBS region. TFBS regions were classified as follows: within 5 kb of the 5′ most exon of a gene, within 5 kb of the 3′ terminal exon, or within a gene, novel or outside of any annotation, and pseudogene/ambiguous (TFBS overlapping or flanking pseudogene annotations, limited to chromosome 22, or TFBS regions falling into more than one of the above categories).

What's wrong with this one?

The three-dimensional rendering of this pie chart is gratuitous, and pie charts are bad anyway, as humans are notoriously poor at comparing areas. The color is gratuitous, too. Any graph that is meaningful only if the numbers are also cited must be viewed as a failure. That the figure is obviously from Microsoft Excel is another embarrassment.

What should have been done?

The authors could have just cited the numbers. Alternatively, a bar plot (without gratuitous 3D) wouldn't be unreasonable.

7. Bell ML, et al. (2007) Spatial and temporal variation in PM2.5 chemical composition in the United States for health effects studies. Environmental Health Perspectives 115:989-995, Figure 3
[The article | The figure | Discussion]

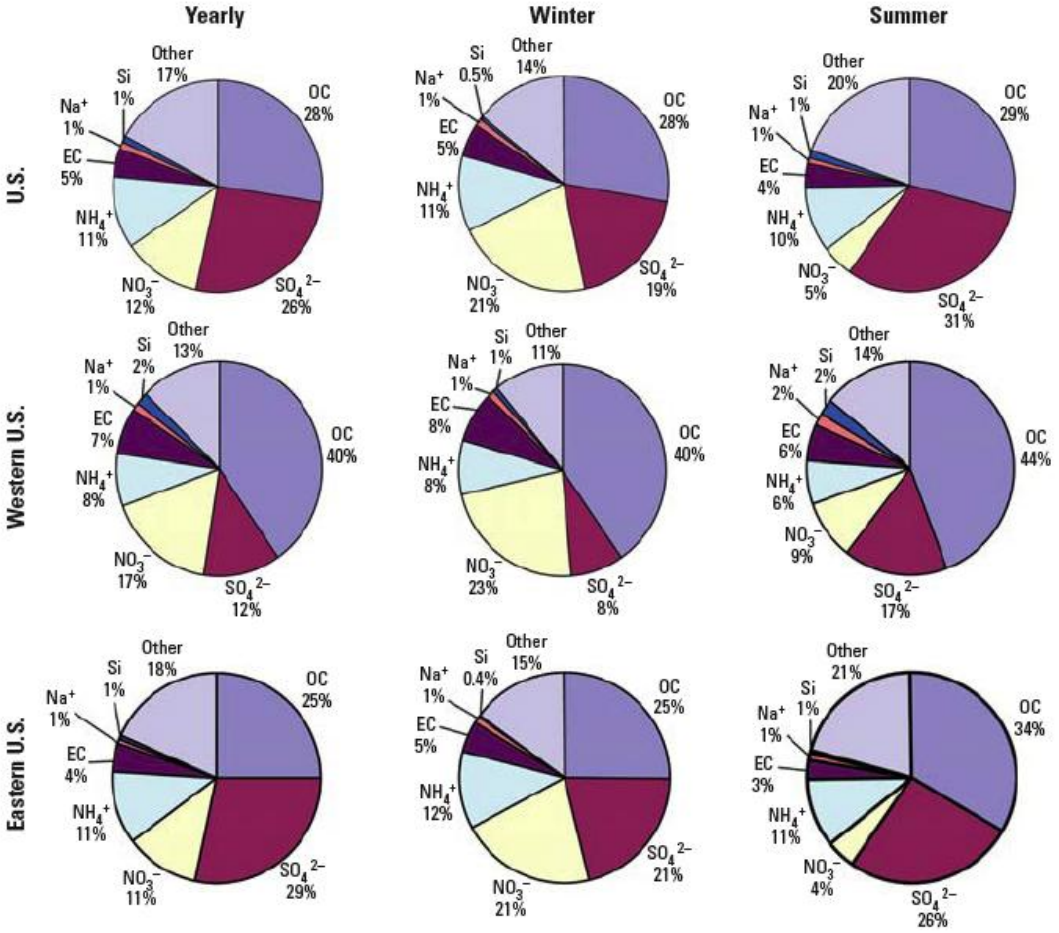


Figure 3. Percent of $PM_{2.5}$ composition by component for yearly, winter, and summer averages, by region.

What's wrong with this one?

What's worse than one piechart? Try nine. This figure is intended to demonstrate that NO_3 and SO_4 change dramatically between winter and summer and that the change is consistent across geographic region. Unfortunately, it takes a couple of minutes of squinting to see this. Piecharts are bad because humans are notoriously poor at comparing areas. This figure is even worse because we are expected to compare areas across what appears to be a bake sale of piecharts.

What should have been done?

An easy fix would be to use barplots instead. However, because all that matters is the change from summer to winter, a couple of simple scatter plots (East and West), with winter and summer on the x- and y- axes, would do the trick.

8. Jorgenson E, et al. (2005) Ethnicity and human genetic linkage maps. American Journal of Human Genetics 76:276-290, Figure 2
[The article | Figure 2a | Figure 2b | Discussion]

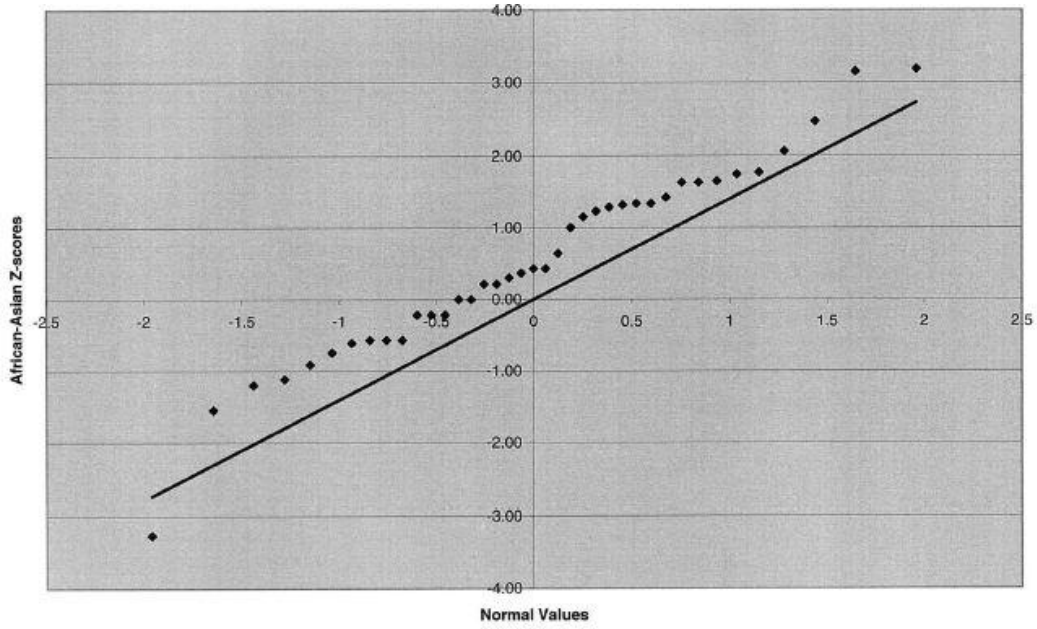
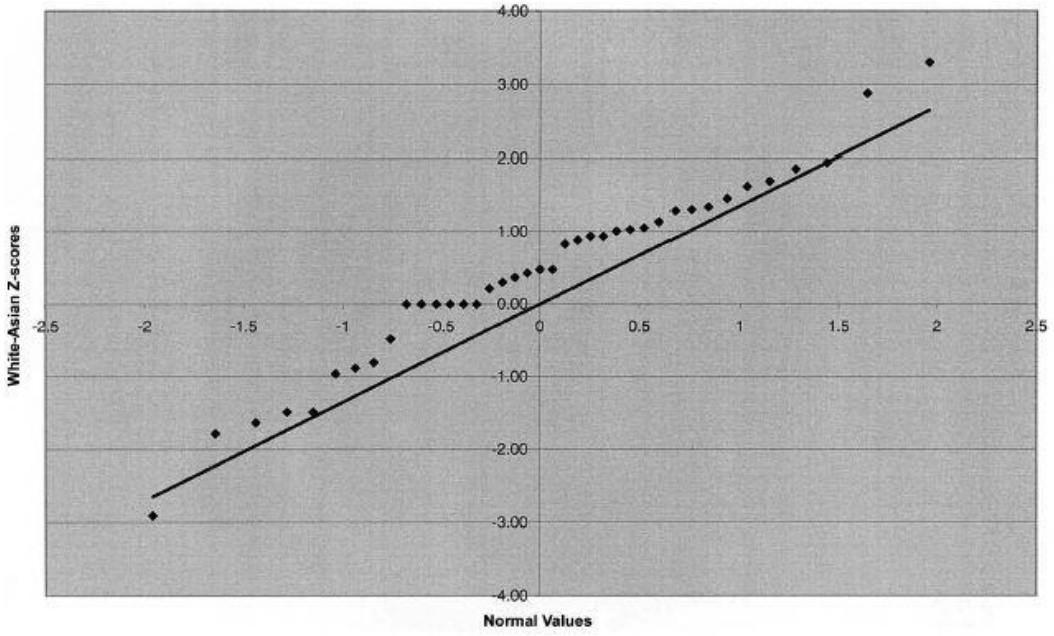


Figure 2 Q-Q plots of Z scores for telomeric interval-length differences. *a*, African Americans versus Asians. *b*, Whites versus Asians.



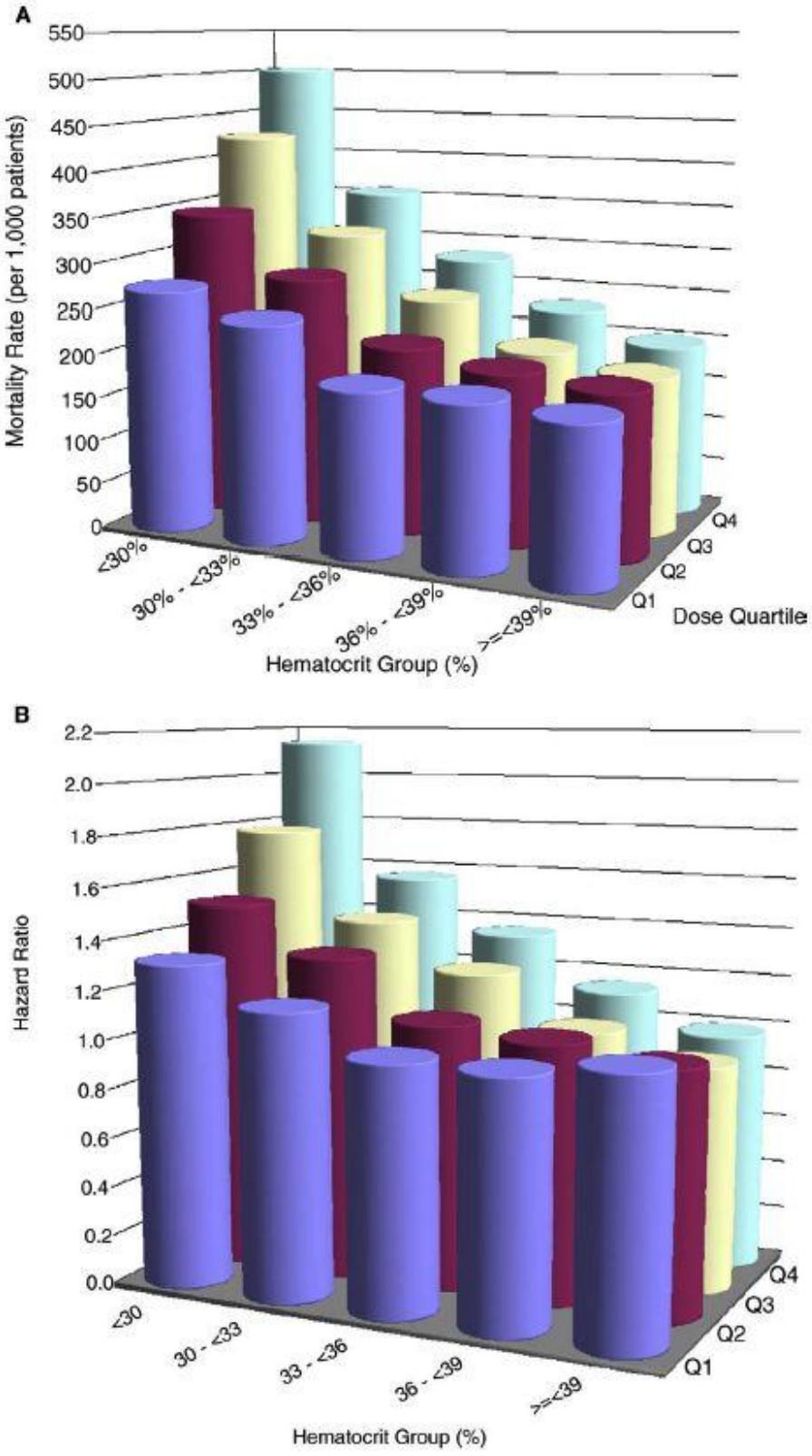
What's wrong with this one?

There's always a lot of wasted space in QQ plots, but these are particularly bad: the figure takes two full pages, and that gray background! I've never seen so much wasted spaced and ink.

What should have been done?

My understanding of this figure is that the authors are comparing three groups of numbers. They do so in a way that is not very natural. I'd prefer dotplots showing the values in the three groups, side-by-side, with a white background, and using one-quarter of the page.

9. Cotter DJ, et al. (2004) Hematocrit was not validated as a surrogate endpoint for survival amoung epoetin-treated hemodialysis patients. Journal of Clinical Epidemiology 57:1086-1095, Figure 2



What's wrong with this one?

The perspective makes it difficult to compare the heights of the cylinders, as the vertical scale changes from front to back. Also, this is a lot of space (and color ink) to convey very little information.

What should have been done?

It must be admitted that it's tricky to fix this figure; one might try four superposed lines.

10. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: Individual and sex-specific variation in recombination. American Journal of Human Genetics 63:861-869, Figure 1
[The article | The figure | Discussion]

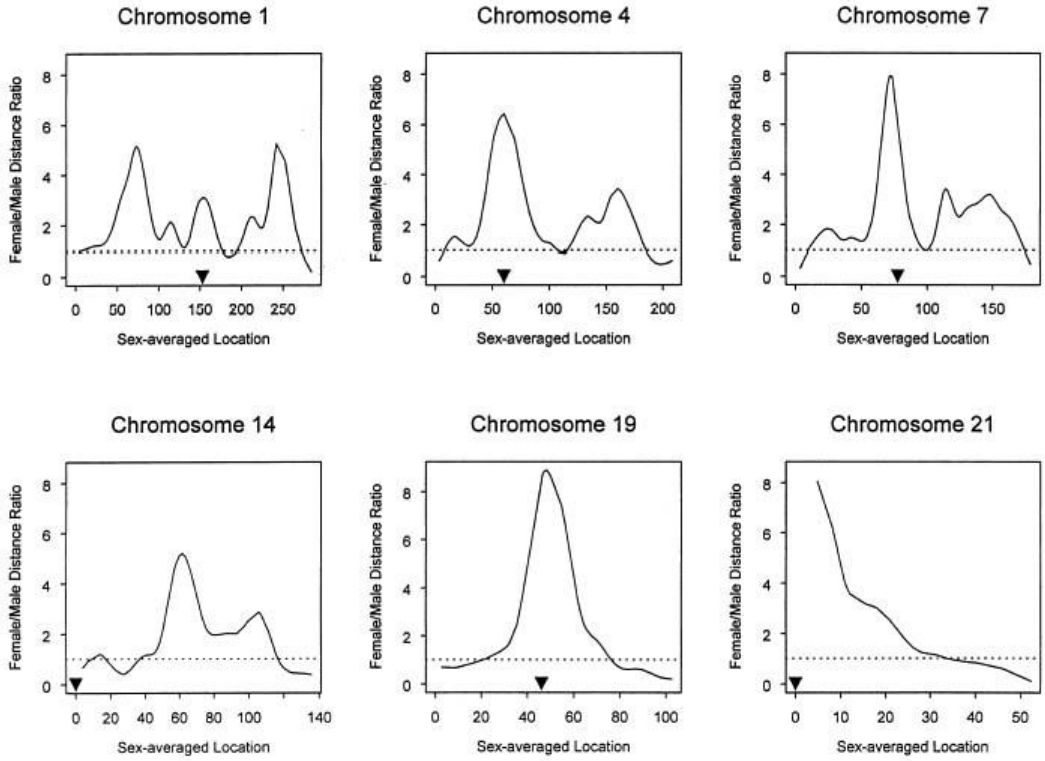


Figure 1 Plots of the female:male genetic-distance ratio against sex-averaged genetic location (in cM) along six selected chromosomes. Approximate locations of the centromeres are indicated by the triangles. The dashed lines correspond to equal female and male distances.

What's wrong with this one?

This is a plot is of ratio of female : male recombination rate. Because ratios below one are sandwiched between one and zero, much greater emphasis is given to ratios above one, when the female recombination rate is greater than the male recombination rate.

What should have been done?

This should have been done on the log scale (preferably log base 2), so that ratios above and below one are given equal emphasis.

Honorable mention: A really bad table

Paik MC (2004) Nonignorable missingness in matched case-control data analyses. *Biometrics* 60:306-314, Table 5

What's wrong with this one?

There are far too many digits (as indicated by the estimated standard errors, which you can't see immediately; take the square roots of the variances---that you have to do that is another problem), and ending zeros were dropped. If the digits in 0.02229 were meaningful, then 0.02100 should be given rather than 0.021.

What should have been done?

One should essentially always cite standard errors rather than variances; they are in the appropriate scale. Use the standard errors to guide you regarding which digits to present; don't include digits that are just noise. Don't drop ending zeros when they are meaningful.

Table 5

Simulation results for using full data, CRs only, and proposed method under four missing mechanisms

	Bias[a]		Variance[b]		95% CI[c]	
Method	$(\hat{\beta}_W)$	$(\hat{\beta}_X)$	$(\hat{\beta}_W)$	$(\hat{\beta}_X)$	$(\hat{\beta}_W)$	$(\hat{\beta}_X)$
			(M.1) $P(R=1)=0.66$			
Full	0.01346	0.02229	0.04008	0.03685	0.955	0.950
Comp	0.03062	−0.003561	0.1149	0.06732	0.960	0.955
Impu	0.01431	0.021	0.04088	0.05169	0.980	0.975
			(M.2) logit $P(R=1)=2Y$			
Full	0.007908	−0.02116	0.03838	0.03624	0.975	0.925
Comp	0.01945	0.07096	0.107	0.06581	0.960	0.950
Impu	0.006966	0.01597	0.04227	0.05226	0.975	0.985
			(M.3) logit $P(R=1)=2X$			
Full	0.007908	−0.02116	0.03838	0.03624	0.975	0.925
Comp	0.01225	0.0589	0.08856	0.06818	0.980	0.975
Impu	0.009563	−0.04699	0.03865	0.04923	0.985	0.970
			(M.4) logit $P(R=1)=X+Y$			
Full	0.01346	0.02229	0.04008	0.03685	0.955	0.950
Comp	0.02404	1.613	0.1102	0.08202	0.955	0.580
Impu	0.01814	0.08289	0.0578	0.06075	0.955	0.970

[a]Bias $= (\hat{\beta}-\beta_0)/\beta_0$.
[b]Simulation variance.
[c]Confidence interval using jackknife standard error.