

写了比没写还糟糕？ 避免User Study中的无效写作



杨馥梦

Games Webinar, 2022年5月5日

我

2009 - 2013	山东大学	计算机科学与技术	学士学位
2013 - 2016	Tufts University	计算机科学	硕士学位
2015 - 2021	Brown University	计算机科学	博士学位

<http://www.fmyang.com/>



User Study

User research focuses on understanding user behaviors, needs, and motivations through interviews, surveys, usability evaluations and other forms of feedback methodologies.

针对某一个问題，以收集并分析人的数据为主，并从中提取知识和经验的一类研究。

这个问题，可以是对某一个现象的假设，也可以是对某一项技术的可用性、有效性等的测试。

可定性（**qualitative**）可定量（**quantitative**）。

我们今天主要讨论定量研究（**quantitative study**）。

无效写作

导师：“…全部重写” “这是啥，这又是啥”

审稿人：“…这段全删了” “我实在看不懂”

如何避免

思路清晰、重点清楚

尽力遵从当前的“业界标准”、熟悉套路

根据具体情况调整

故事纯属虚构
经验真的有用

<https://github.com/Fumeng-Yang/Games-Webinar-0505>

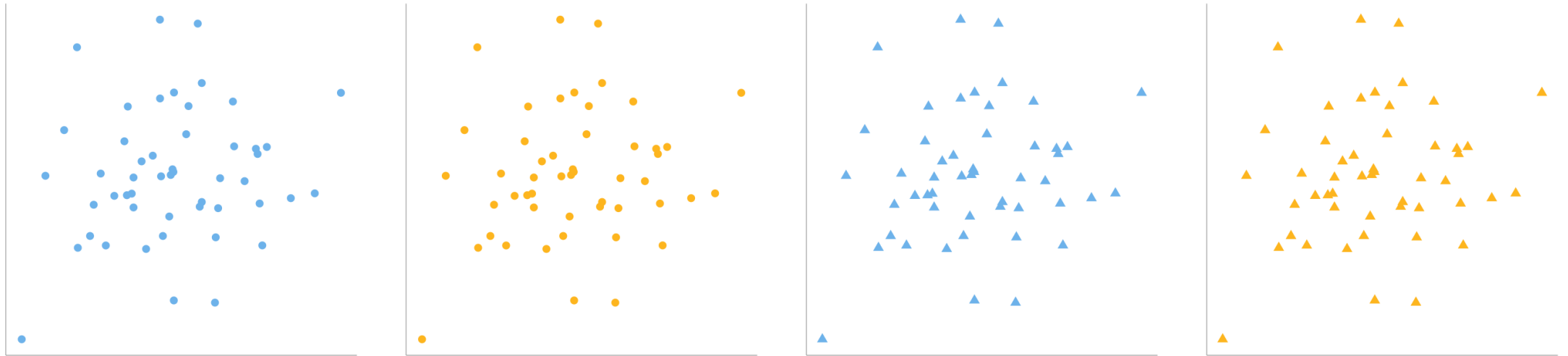
我有一个大胆的想法……

我觉得散点图的**颜色**和**点的形状**对大家看到这个图的时候的**心情数值**有不同影响!

我挑了两种颜色（**蓝**、**黄**），两种形状（**▲**，**●**）。

招了50个参与者，每个人看4张散点图。他们从0（悲伤）到10（快乐）评价自己的心情数值。

为简化问题，我们考虑连续而非离散的心情数值。



(一些实验设计问题，比如repeated trials, within-subjects, between-subjects 此处略过)

Study

Research question: 散点图的颜色和点的形状对心情数值影响不同

Variables: 颜色（蓝、黄），形状（▲，●）。

Participants: 50人

Experiment: 每个人看4张散点图，并从0（悲伤）到10（快乐）评价自己的心情数值。

color	mark	participant	answer
blue	circle	P1	3.1164795
blue	triangle	P1	4.0272738
yellow	circle	P1	8.8108288
yellow	triangle	P1	3.1262111
blue	circle	P10	4.8314713
blue	triangle	P10	3.9486304
yellow	circle	P10	6.6990914

套路

Introduction		非今日的重点，具体问题具体分析
Related work		
Study design		简述
Analyses		重点，尤其画图
Results		
Discussion		非今日的重点，具体问题具体分析
Conclusion		

Introduction & Related Work

可视化中的颜色理论

可视化中使用不同的形状对认知感知的影响

情绪/压力影响因素

Study Design

套路（具体可能调整顺序或者增减）

Research question - “我觉得…”

Visualization - “散点图”

Task - “评价心情数值”

Generating stimuli - 主要是背后的data是怎么生成的

Experimental design - “每个参与者看了所有20张图”

Procedure - “实验流程”

Apparatus - 比如用了VR设备（并没有所以不写）

Measures - “心情数值”

Participants - 如何招到参与者、性别年龄等

Data - 50人乘以4 = 200个回答

这个实验比较简单，可考虑合并

这个实验比较简单，可考虑合并

一些细节要遵循“业界标准”

Analyses and Results

如果Analyses不是很复杂（比如significance tests），推荐合并成一个Section。许多审稿人也愿意一起看。
如果Analyses很复杂（比如复杂的model），推荐分开。

套路

Research questions (读到这里大家已经忘了研究的问题) - 散点图的颜色、形状对用户心情的影响

Hypotheses - 与Research questions二择一。个人倾向RQ，自由度大，避免和假设相反。

接下来一般分子章节写不同的侧面：

Color

Shape

Interaction effects

这个案例很简单，我们可以合并写。

Analyses and Results

警惕

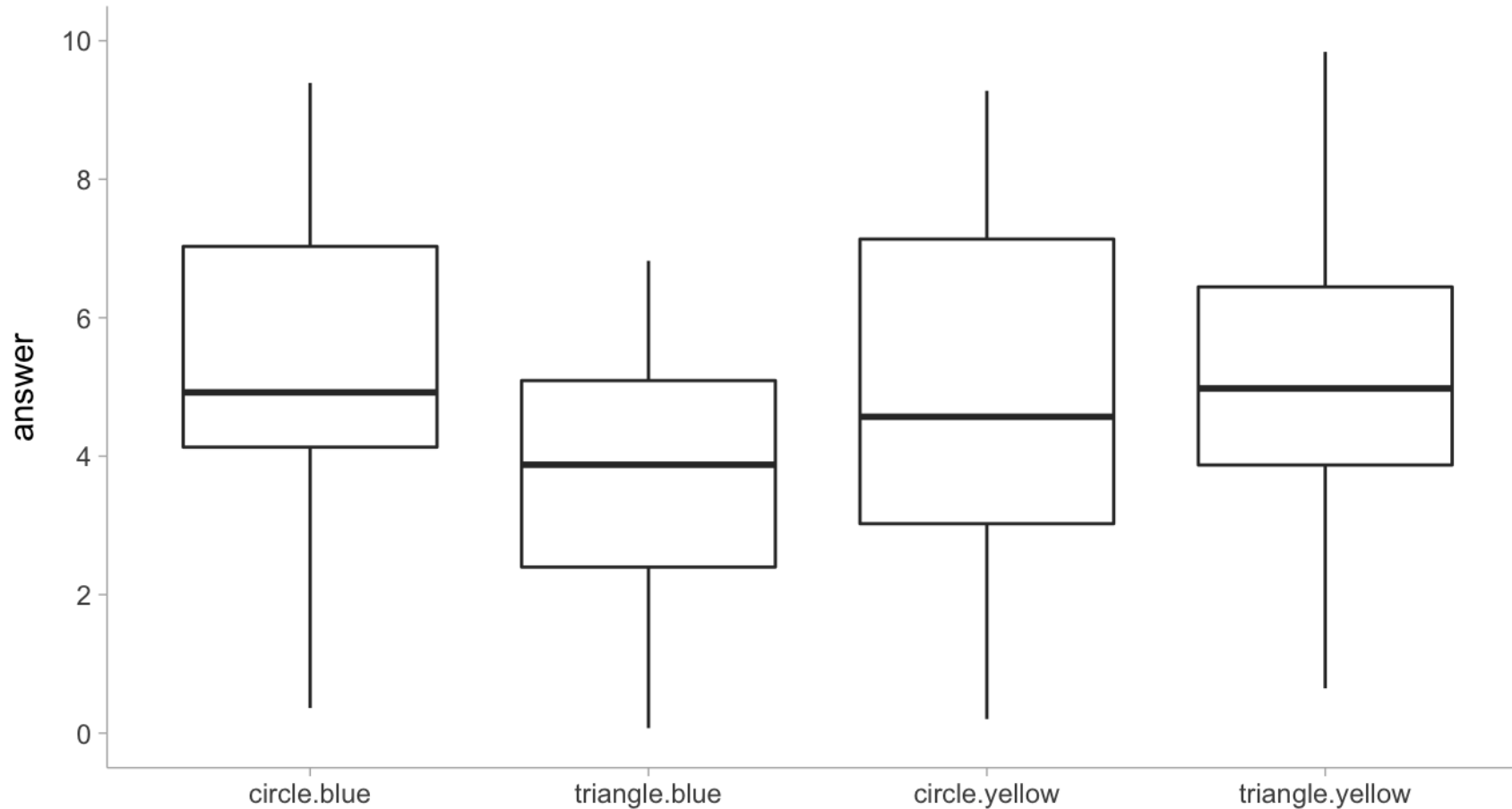
不要把Results埋在一堆Analyses之间。

方法写清楚很重要，但大家更关心结果有没有意思。

不一定所有人对方法都很熟悉。

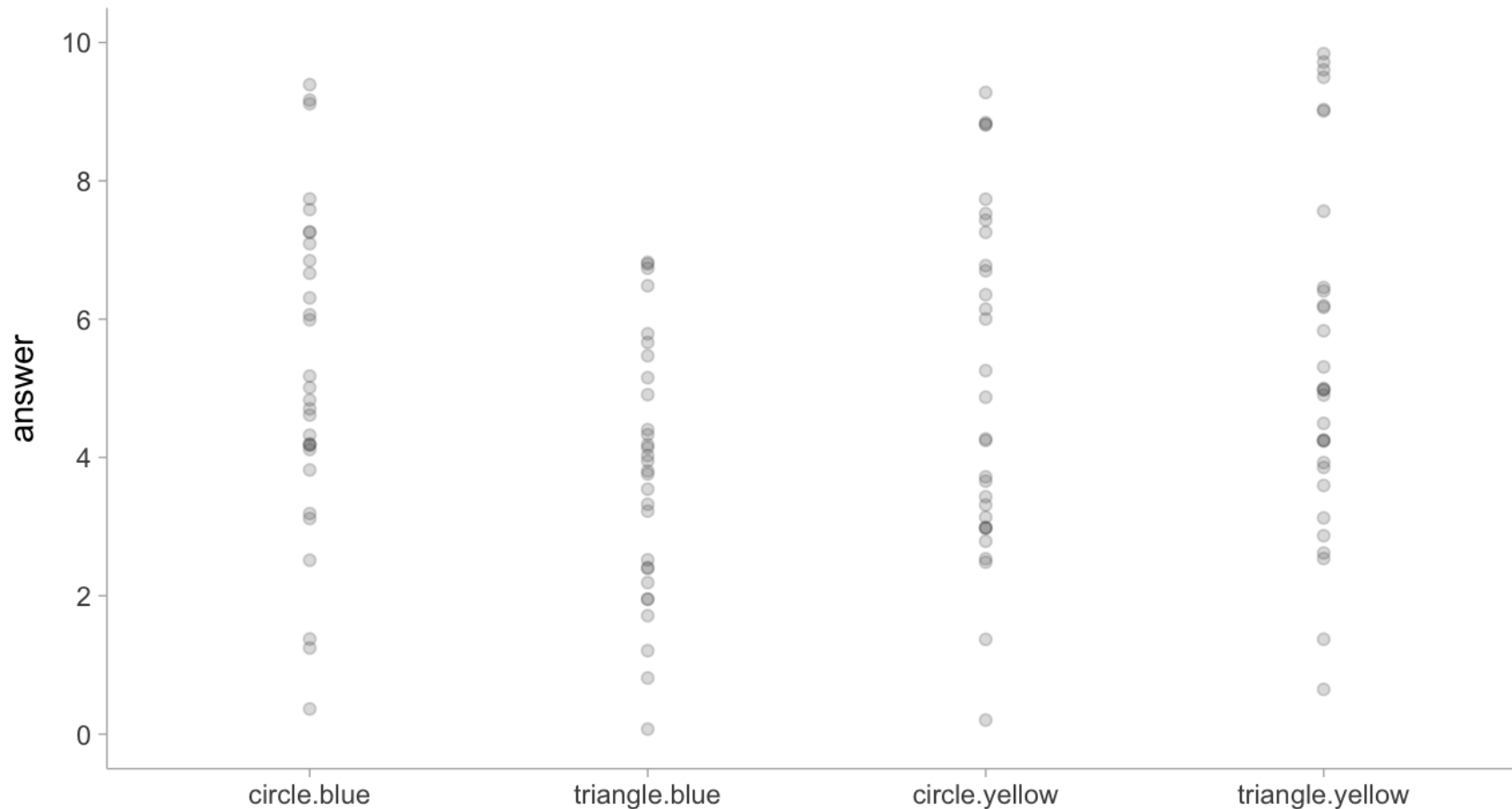
你省了事情，审稿人看得吐血。

Analyses and Results



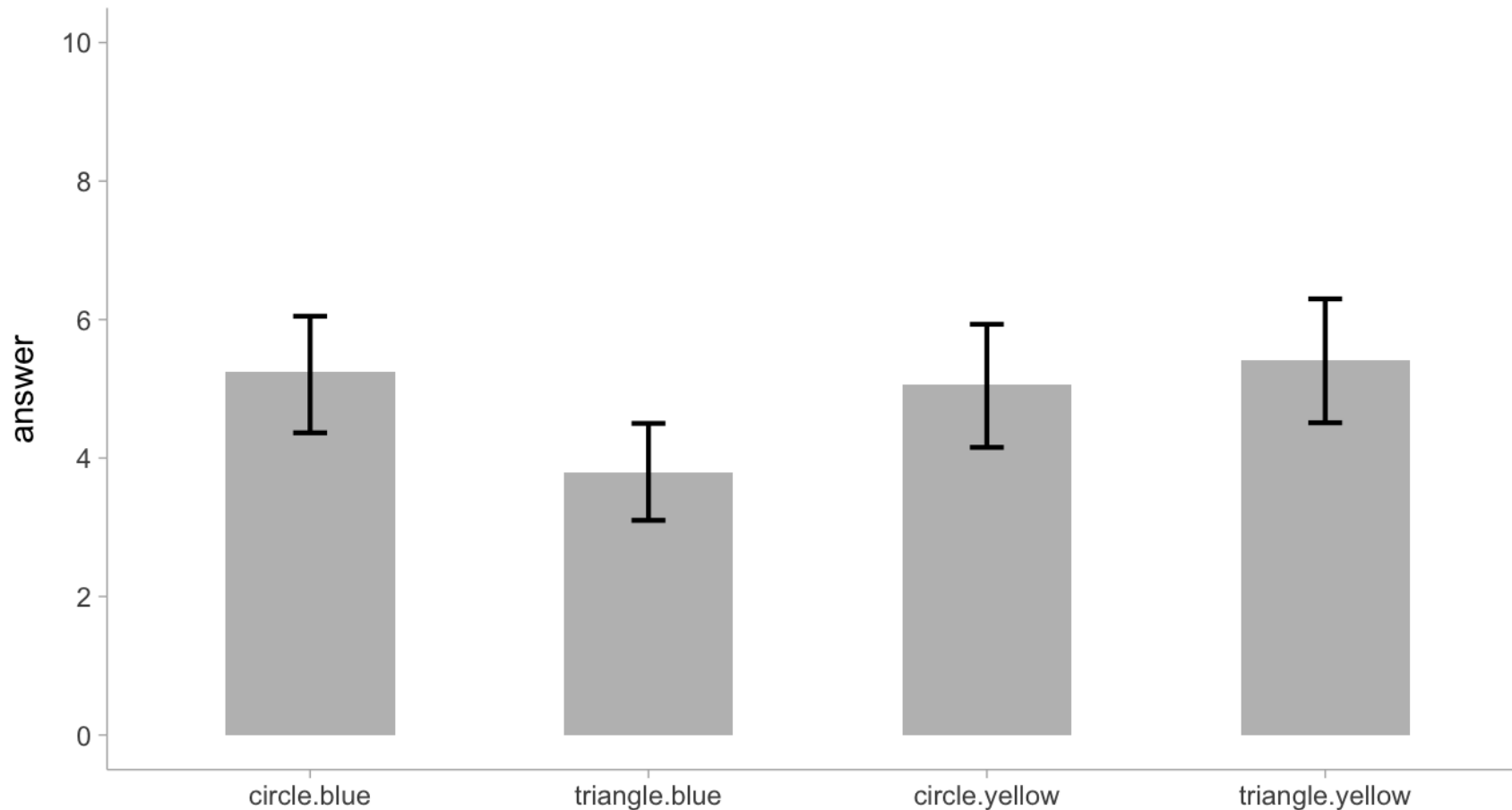
呃，这告诉我什么呢？

Analyses and Results



要不都画出来?
好像更看不清楚了?

Analyses and Results

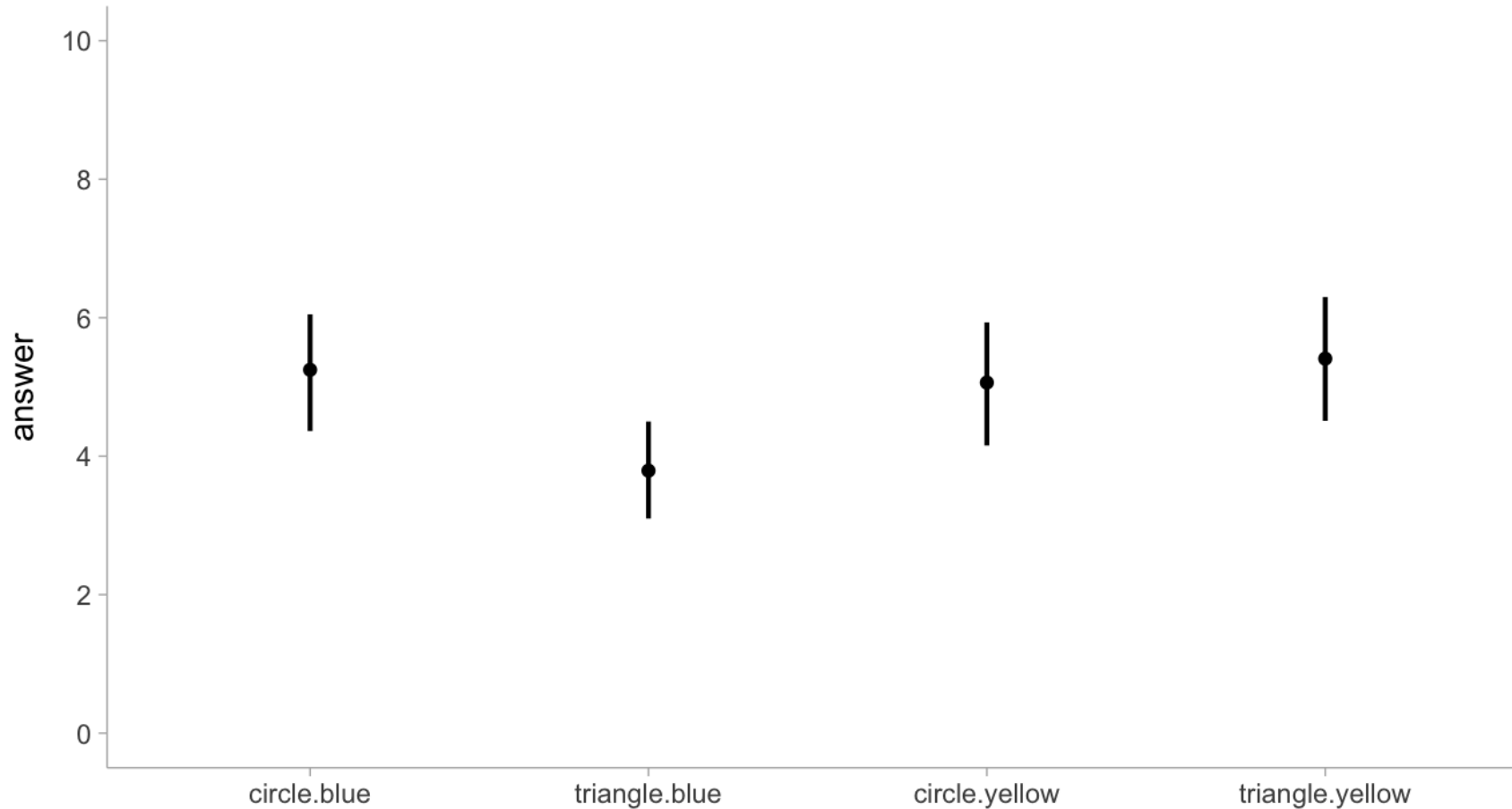


“一生挚爱”：

Bar chart + error bar

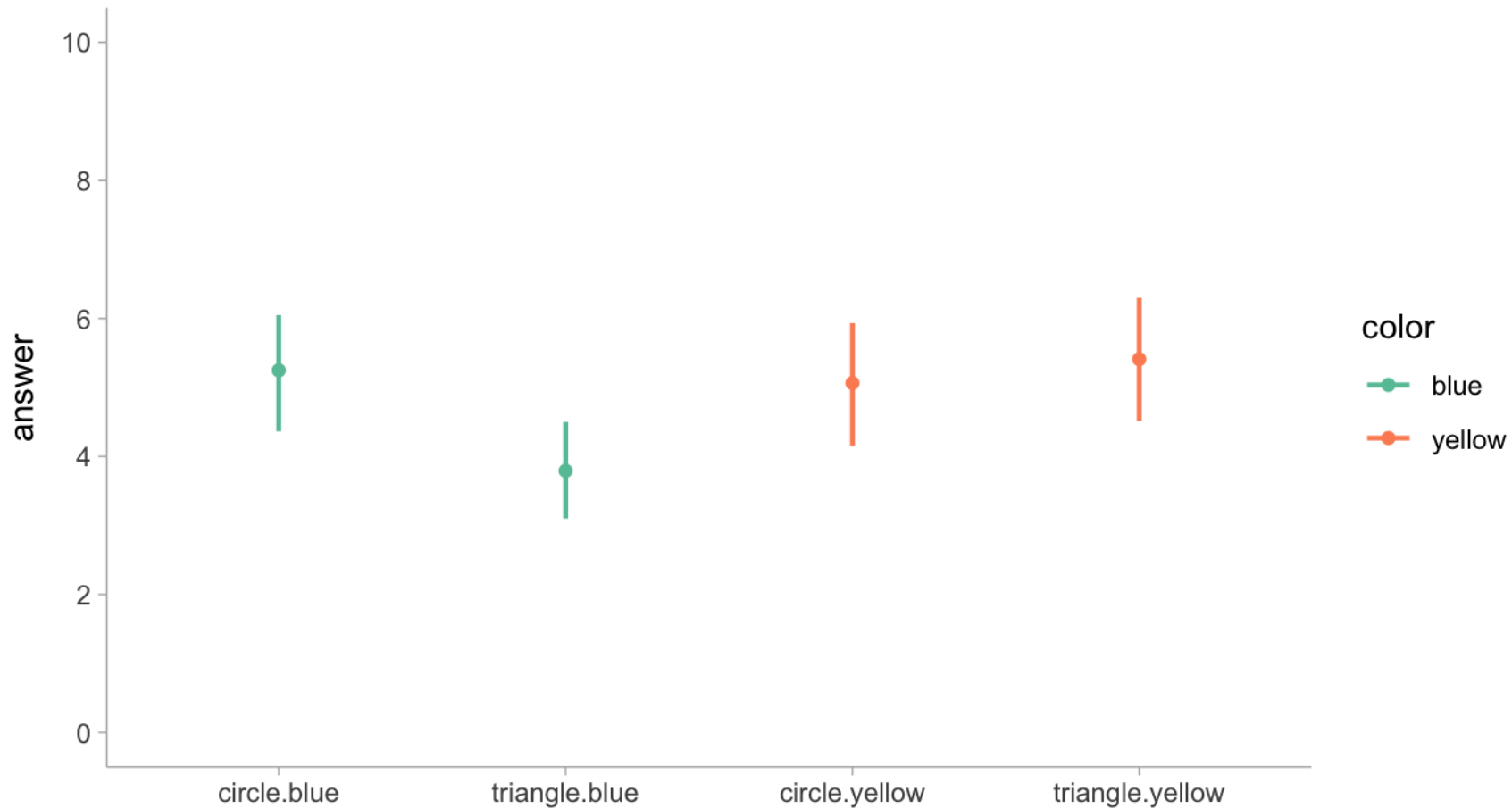
好像有点太老套？注意力都被bars给吸引了？

Analyses and Results



mean + error bar: 非常标准，但是总觉得好像缺点了什么，而且太朴实了

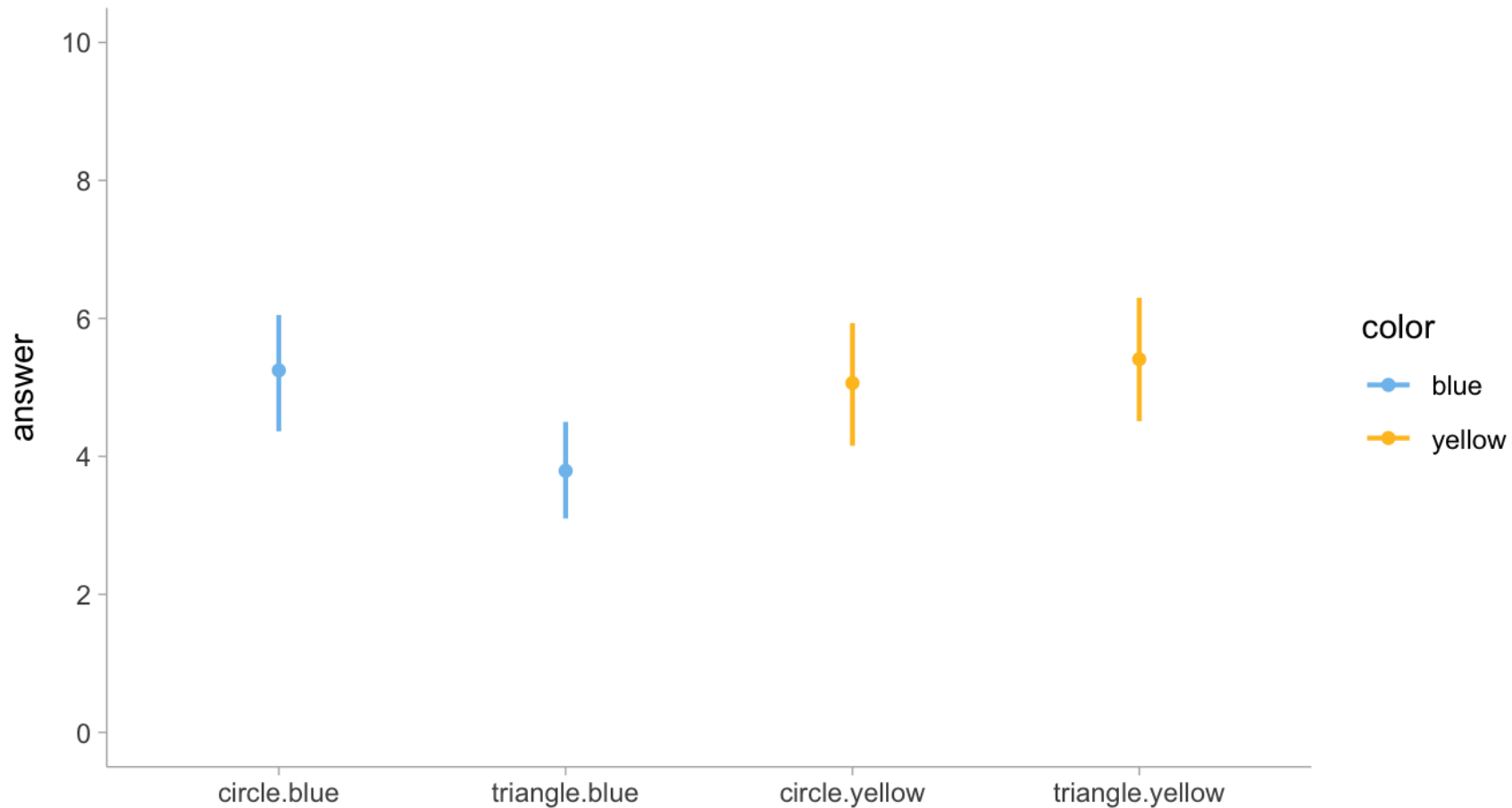
Analyses and Results



加点颜色吧!

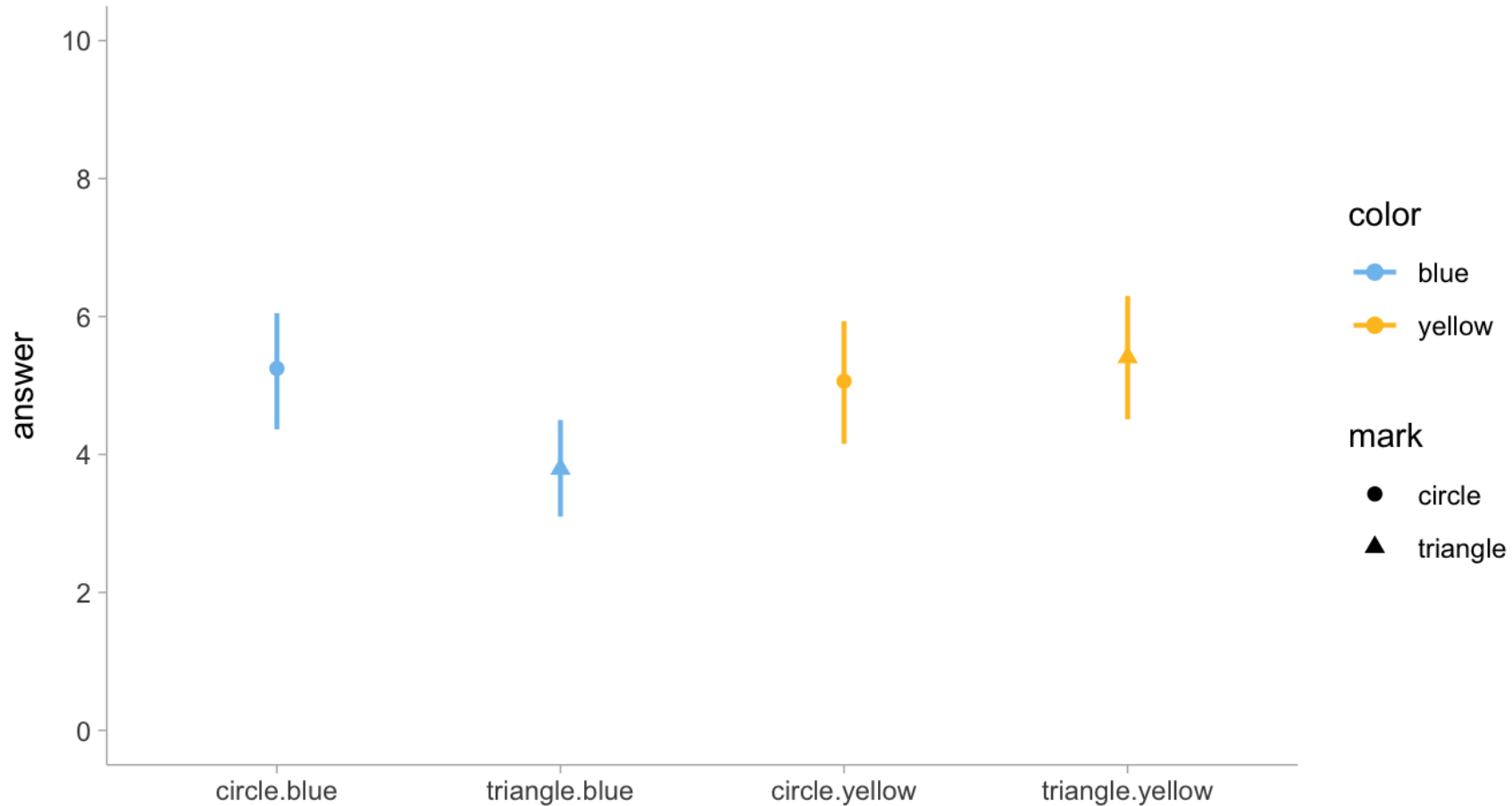
呃, 好像有哪里不对?

Analyses and Results



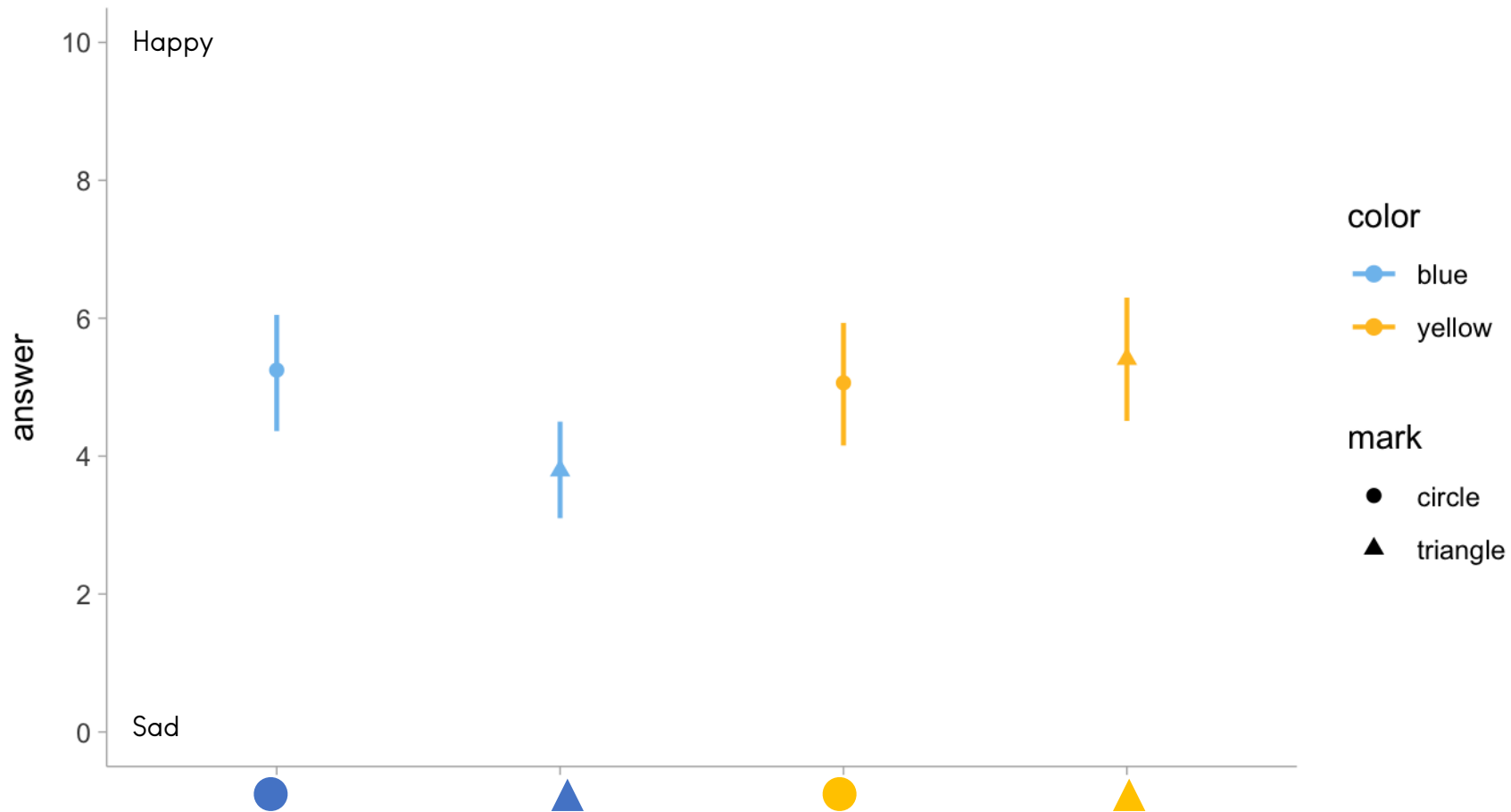
这次对了!

Analyses and Results



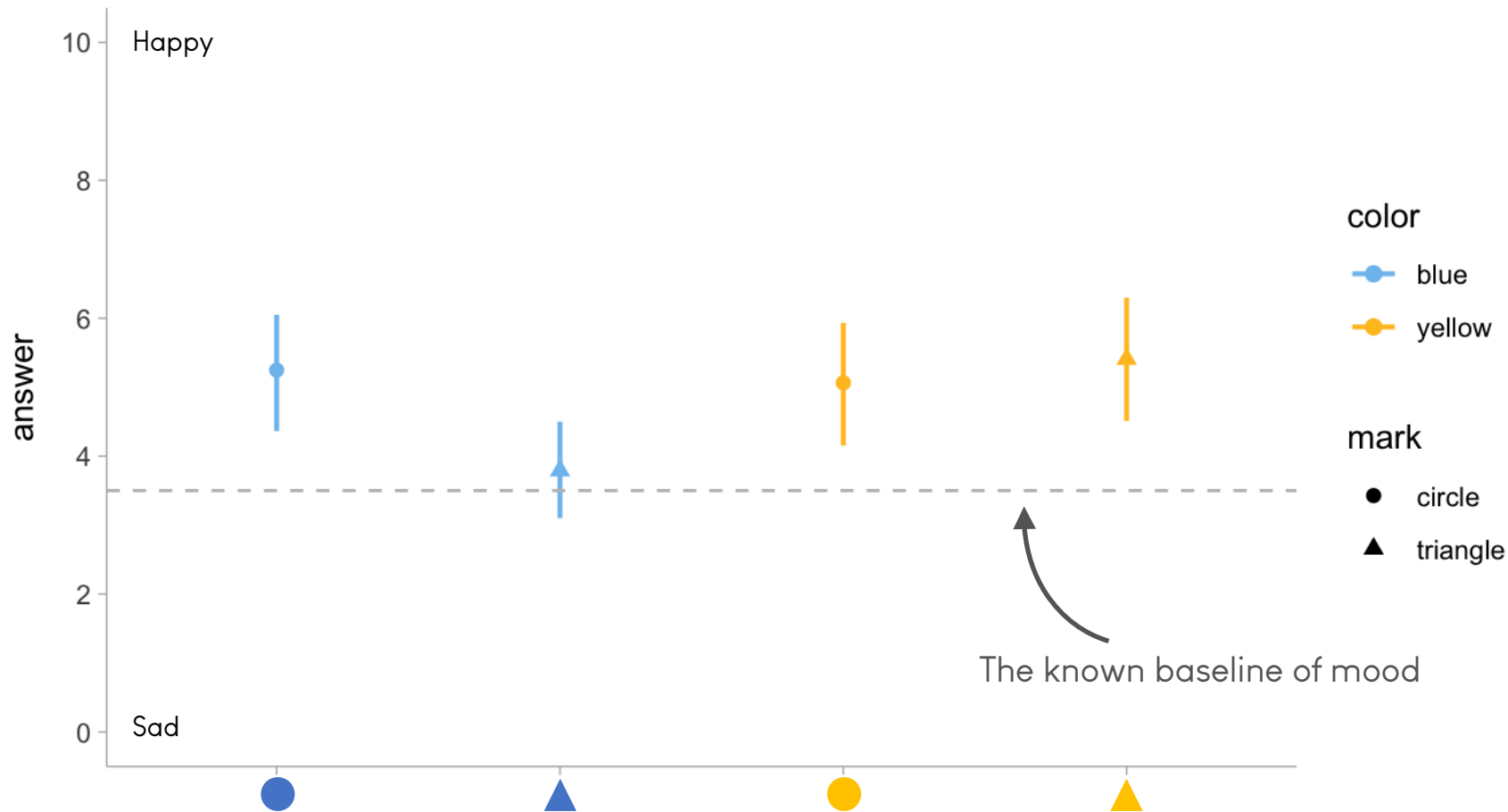
我们可以继续增加表意

Analyses and Results



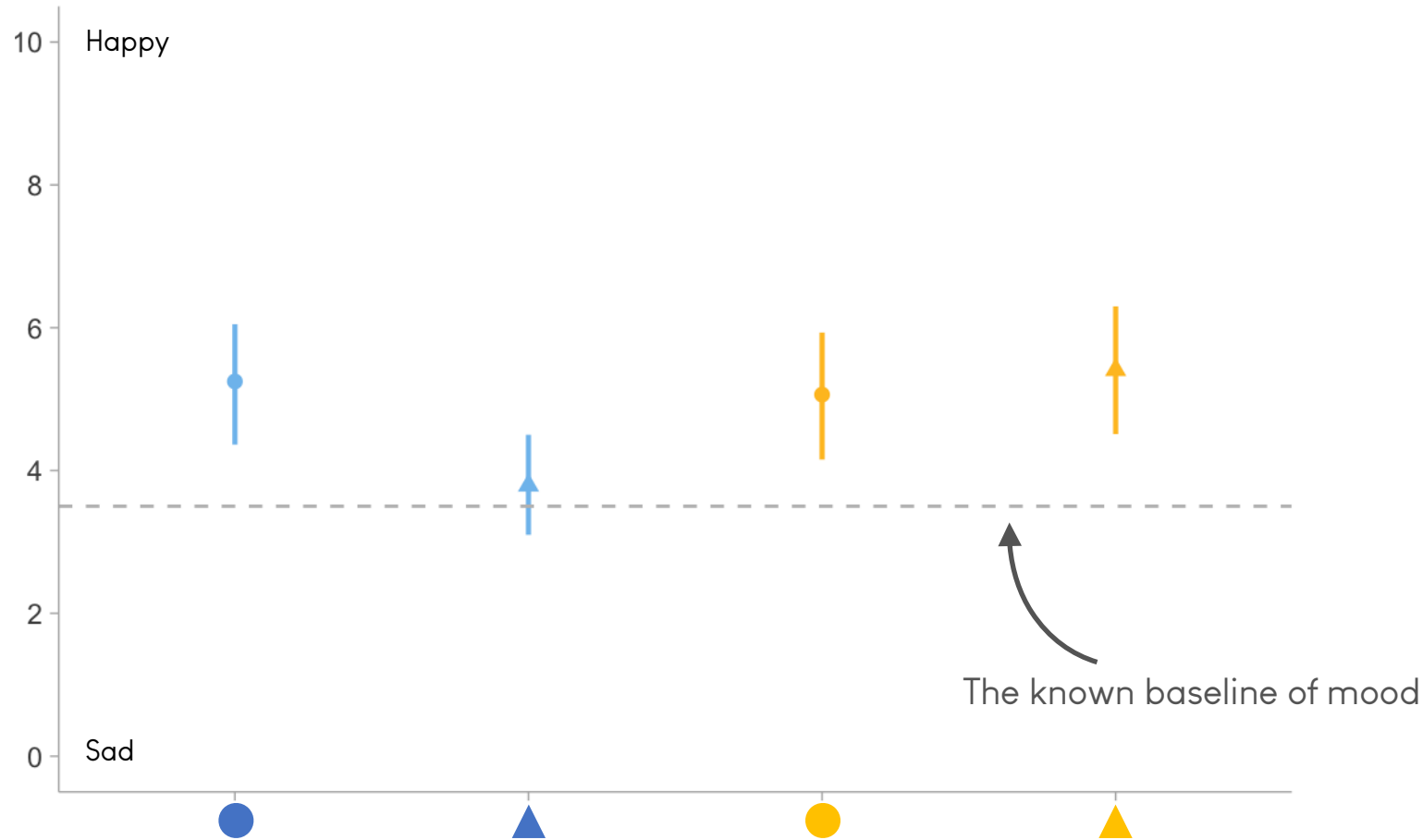
继续

Analyses and Results



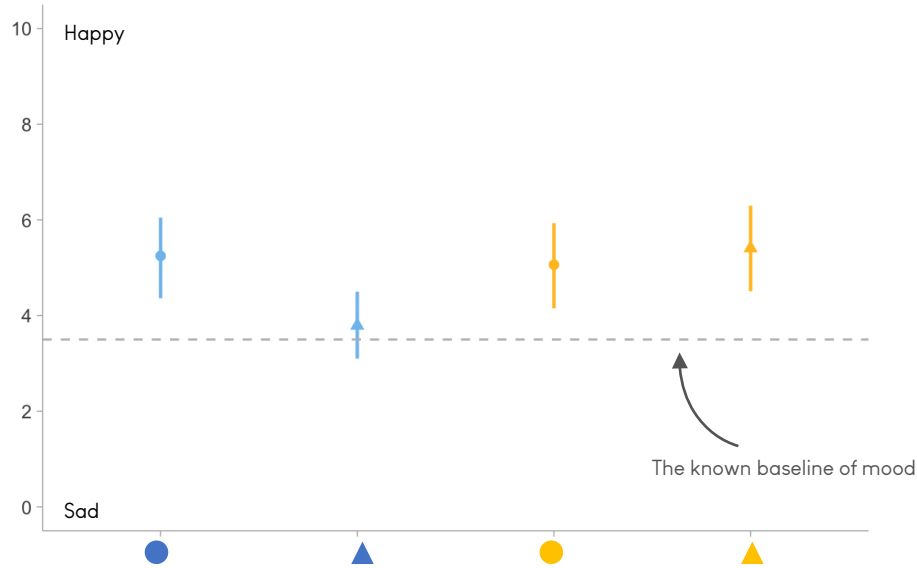
增加一个baseline, 比如是
已知的大家的基准心情

Analyses and Results

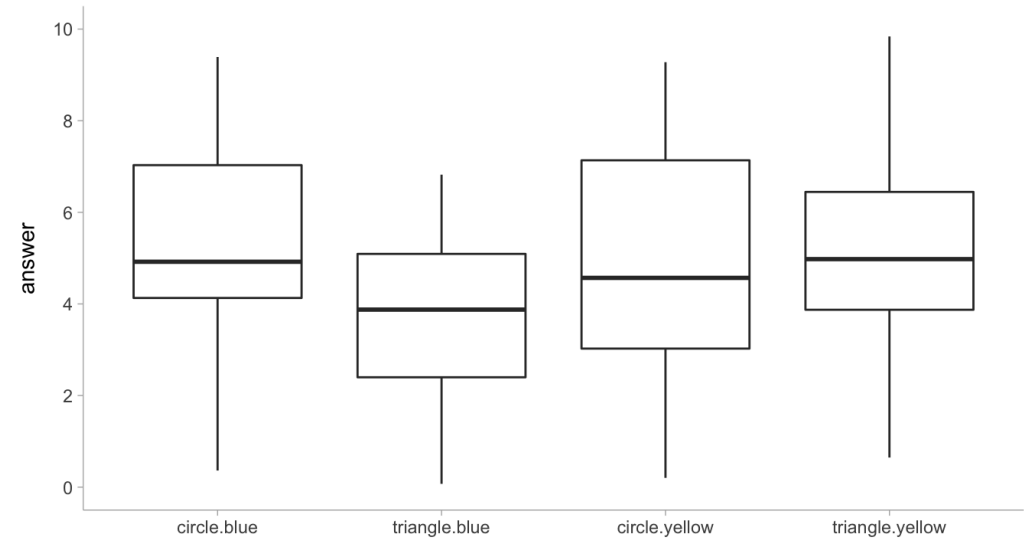


信息够了，开始简化
图例其实没什么必要

Analyses and Results



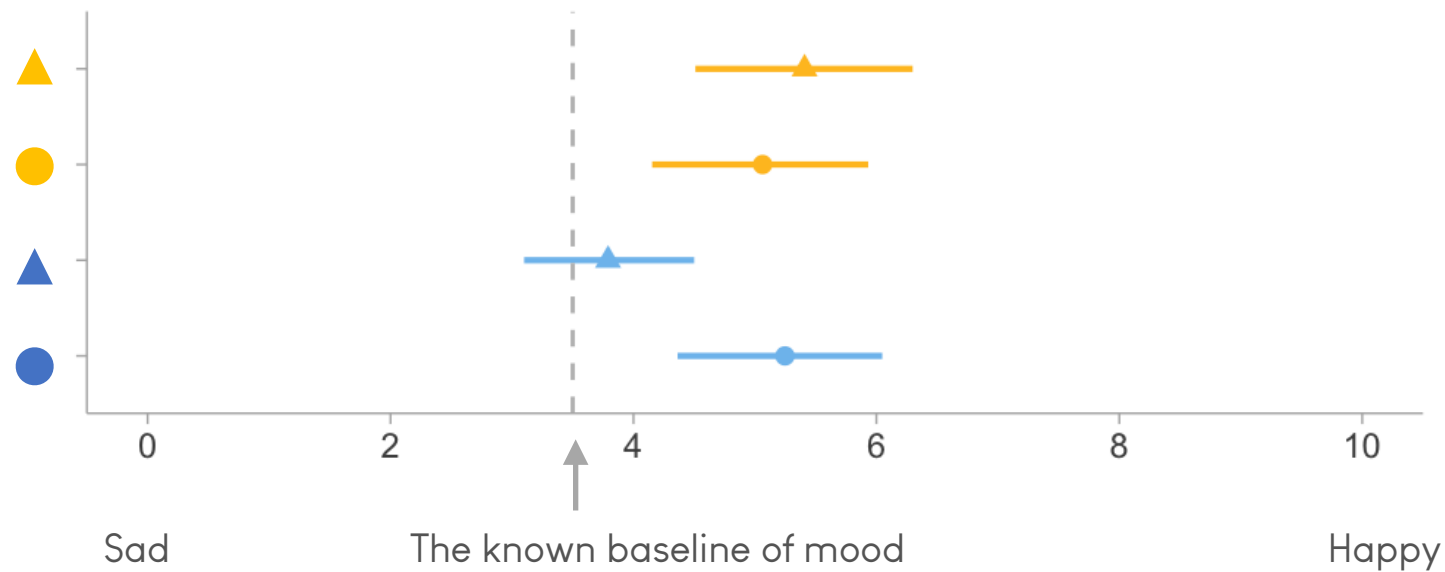
和



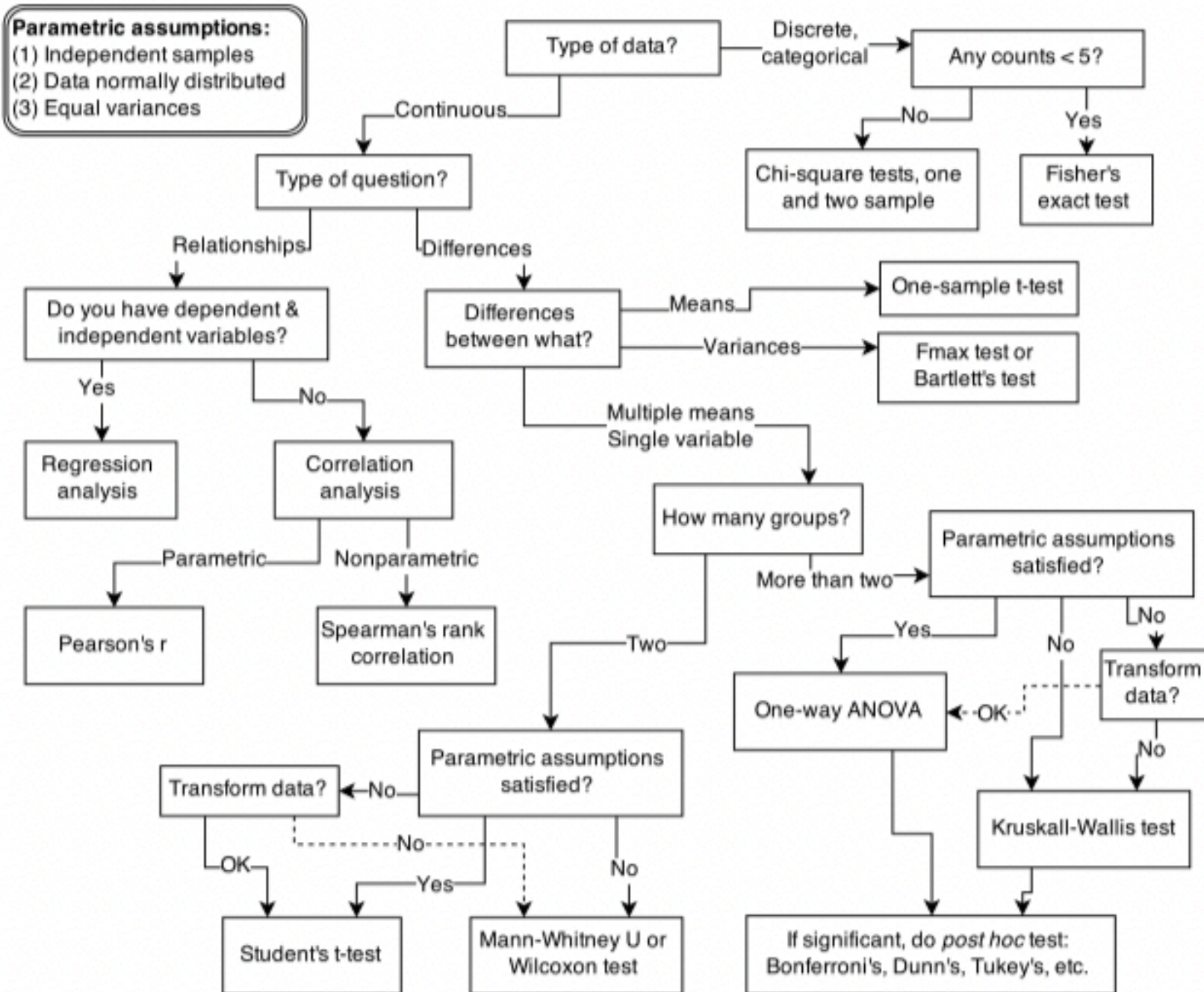
是不是表意和审美方面都好了很多？

可以直接从mean+error bar上推断significance, boxplot虽然经典但做不到这一点。

Analyses and Results



但是因为paper都是纵向长方形的，其实纵向排版挺占地方的，我们尝试横向排版。
满意！下面开始做statistics。



我们用significance test。

其他可考虑：只report置信区间或者Bayesian。

<https://hal.inria.fr/hal-01377894>

Analyses and Results

ANOVA + PostHoc: ANOVA测一下颜色和点的形状有没有区别, PostHoc测一下到底那些不一样

\$color

	diff	lwr.ci	upr.ci	pval
yellow-blue	0.716643	-0.1180155	1.551302	0.0917 .

\$mark

	diff	lwr.ci	upr.ci	pval
triangle-circle	-0.5554196	-1.390078	0.2792389	0.1901

	diff	lwr.ci	upr.ci	pval
yellow:circle-blue:circle	-0.1844695	-1.78419472	1.4152557	1.0000
blue:triangle-blue:circle	-1.4565321	-3.05625731	0.1431931	0.0962 .
yellow:triangle-blue:circle	0.1612234	-1.43850178	1.7609486	1.0000
blue:triangle-yellow:circle	-1.2720626	-2.87178778	0.3276626	0.2095
yellow:triangle-yellow:circle	0.3456929	-1.25403225	1.9454181	1.0000
yellow:triangle-blue:triangle	1.6177555	0.01803034	3.2174807	0.0459 *

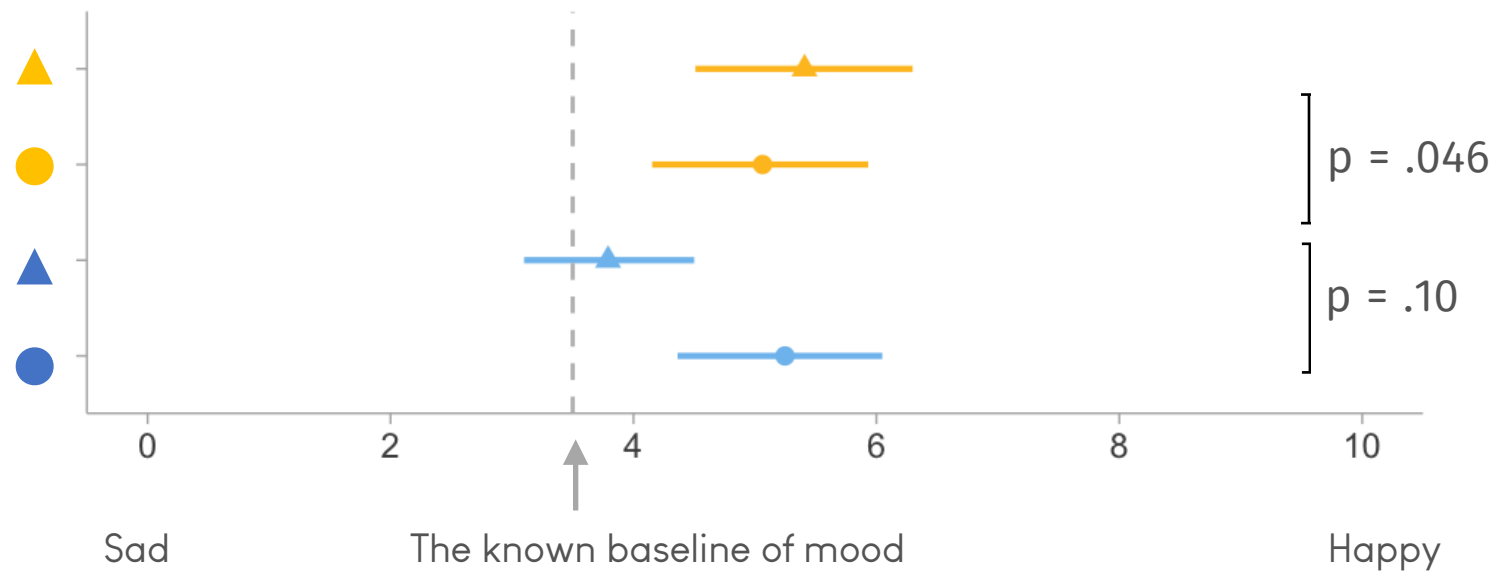
Analyses and Results

“业界标准” : <https://apastyle.apa.org/instructional-aids/numbers-statistics-guide.pdf>

$F(\dots) = \dots, p = \dots, \text{mean} = \dots, \text{sd} = \dots, \dots, F(\dots) = \dots, p = \dots, \text{mean} = \dots, \text{sd} = \dots, \dots, F(\dots) = \dots, p = \dots, \text{mean} = \dots, \text{sd} = \dots, F(\dots) = \dots, p = \dots, \text{mean} = \dots, \text{sd} = \dots, \dots$

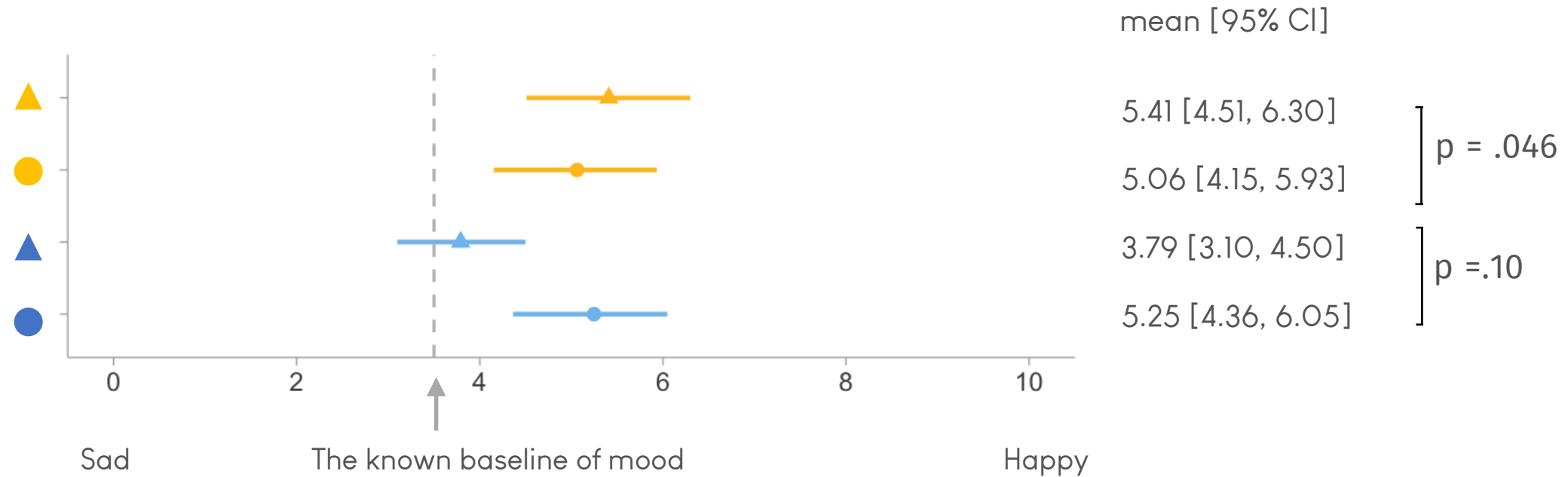
眼花了…

Analyses and Results



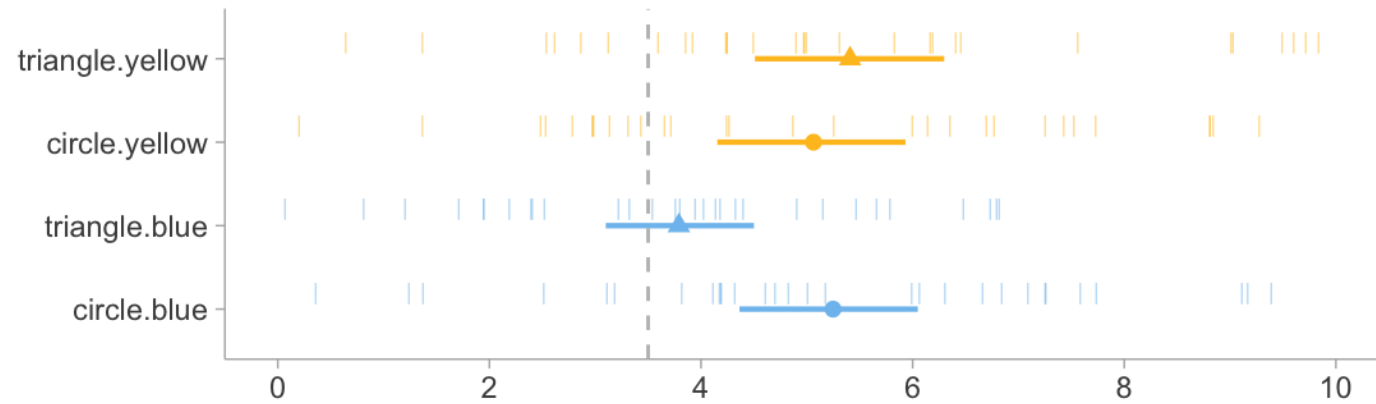
我们可以把significant的结果标在图上。

Analyses and Results

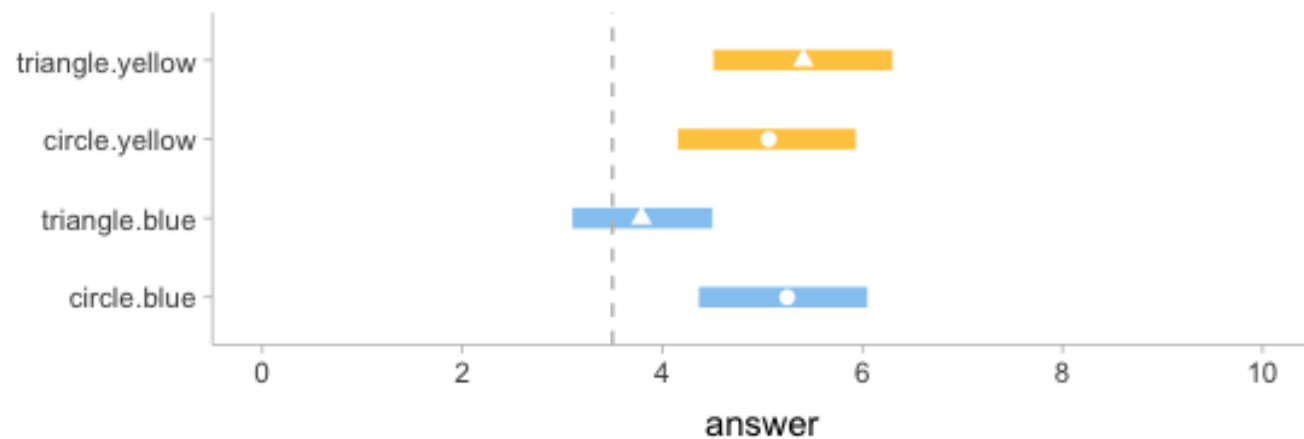


有些统计方法可以不用significance test, 比如靠confidence interval。
这时候我们考虑直接把confidence interval 标在图旁边。
差不多很满意了!

Analyses and Results



其他想法



Analyses and Results

别忘了写error bar的含义：是标准差还是置信区间。

用最直白的语言解释结果：平均来讲，颜色和点的形状在影响参与者心情上没有显著区别。但黄色三角形相比于蓝色三角形显著改善参与者的心情。并且黄色三角形、黄色圆形、蓝色圆形相比于基准线，均能改善参与者的心情。

https://judyrobertson.typepad.com/files/chi2012_submission_final.pdf

Discussion, Conclusion

(略)

提醒

统计方法选择、结果报告尽量遵循“业界标准”，可以找已经发表的论文参考

目前趋势是提高transparency, 在审稿过程中就提交code和data

整理好，做好说明，并且最好在最初就想着怎么让大家明白code和data

千万不要 “the data is available at [withhold for review]” （我觉得比不提还糟糕）

但是还是手稿质量第一，补充材料第二

很多步骤是需要提前考虑好的：是否和怎么去掉outliers, 用什么tests

提醒

虽然套路得人心，根据具体情况要做出调整

不要觉得写完就是目的，**交流**、**表达**、**阐述**我们的研究才是写paper目的

我们花几个月甚至一年做研究、数周写论文，审稿人和其他读者充其量只能看几小时

一定要简洁易懂、逻辑通顺

不要把所有做了工作都写进去

就算没有页数限制，多看一页也更费时间

审稿人看不明白时其实也很痛苦

谢谢!

