



探索性数据分析

课程名称: 跨媒体数据可视化

学生姓名: 宁若汐

专 业: 英语语言文学

学 号: 3190105959

浙江大学 2021.12.15

探索性数据分析 3——宠物小精灵数据挖掘

对所有游戏玩家来说，这是一个包含来自七代 802 个 Pokémon 的信息数据集。试着回答以下几个问题：

1. 哪一代宠物小精灵最强？哪代最弱？

结论：第七代对特殊技能抗性最强，第四代基础属性最强；第二代在两方面都最弱。

分析：

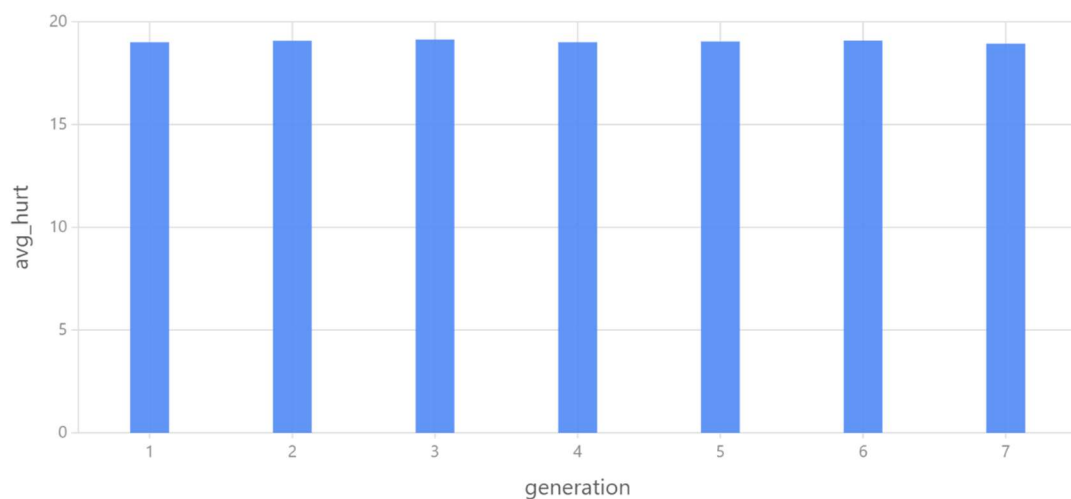
该问题要求探索“代”与“能力”之间的关系。可以看到数据给出了“代（generation）”的标签，但能力的判断标准需要自定义。阅读数据，结合宝可梦游戏经验，我们取“基础属性”和“抗性”两个指标评判宝可梦的能力。

“基础属性”处理方法：游戏的基础属性定义为攻击（attack）、特殊攻击（sp_attack）、防御（defense）、特殊防御（sp_defense）、速度（speed）、体力（HP）六项指标之和。数据中已给出每个宝可梦该六项指标之和数据（base_total），我们可以直接使用这组数据作为“基础属性”指标。

“抗性”处理方法：游戏对“抗性”的定义是，假设某宝可梦在其它属性相同情况下受到 18 种不同属性攻击，数据含义如下表所示：

数据	0	0.25	0.5	1	2	4
含义	无效攻击	4 倍抵抗	收效甚微	正常效果	效果拔群	4 倍伤害

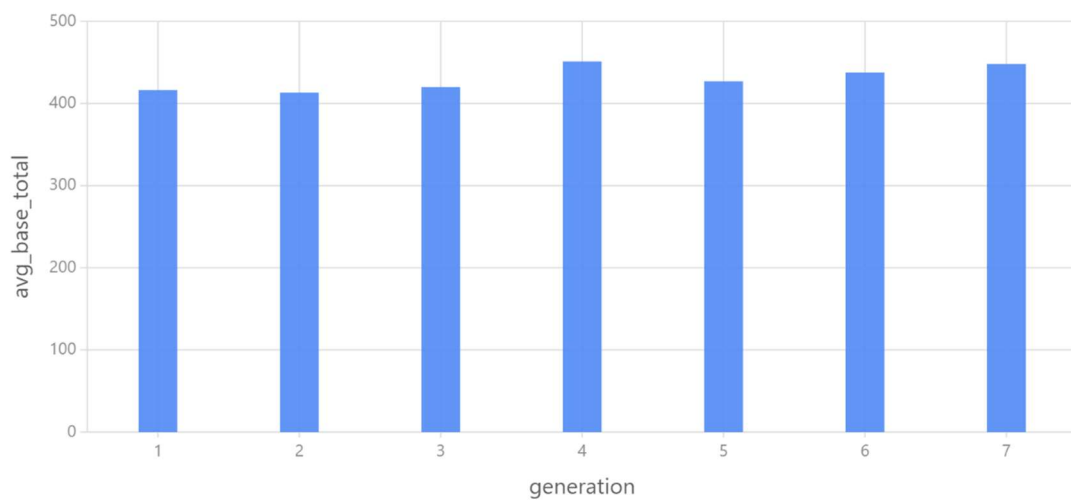
即数值越高，代表宝可梦对该属性攻击防御效果越差，对该属性抗性越低。由于所有抗性数据取值范围相同，可以不加权对 18 属性抗性求和得出每个宝可梦的抗性总和。



天枢生成的平均伤害与代的关系图。读图可知：

代	1	2	3	4	5	6	7
平均伤害	19.00	19.08	19.13	19.00	19.04	19.09	18.93

因为受伤害越低表示宝可梦抗性越高，则能力越强。由第七代伤害最低可知，第七代宝可梦在抗性方面最强；而第三代宝可梦伤害最高，则抗性方面最弱。



	1	2	3	4	5	6	7
	416.25	413.18	420.02	451.18	427.04	437.31	448.16

由于基础属性越高，则宝可梦能力越强，可知第四代宝可梦在基础属性方面最强，第二代在基础属性方面最弱。

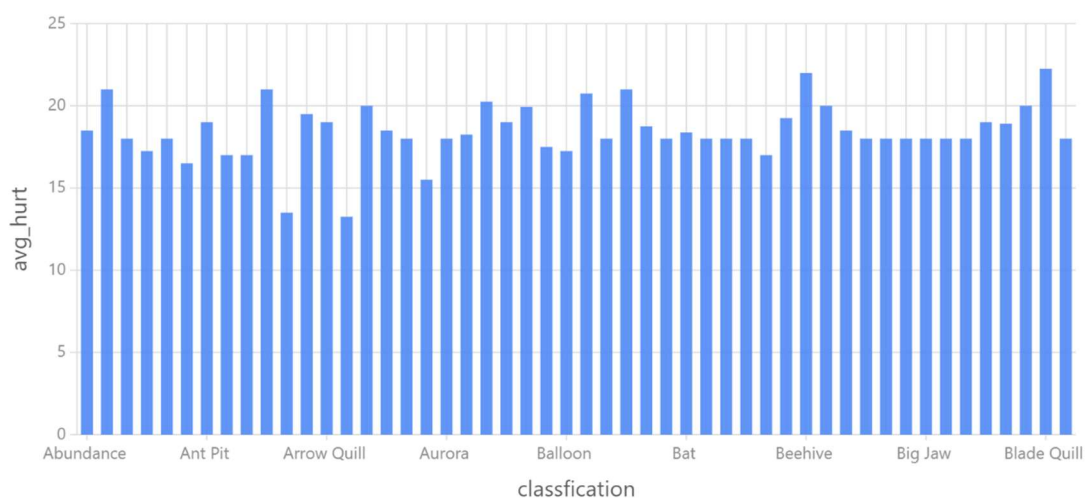
2. 哪种类型神奇宝贝最强？哪种最弱？

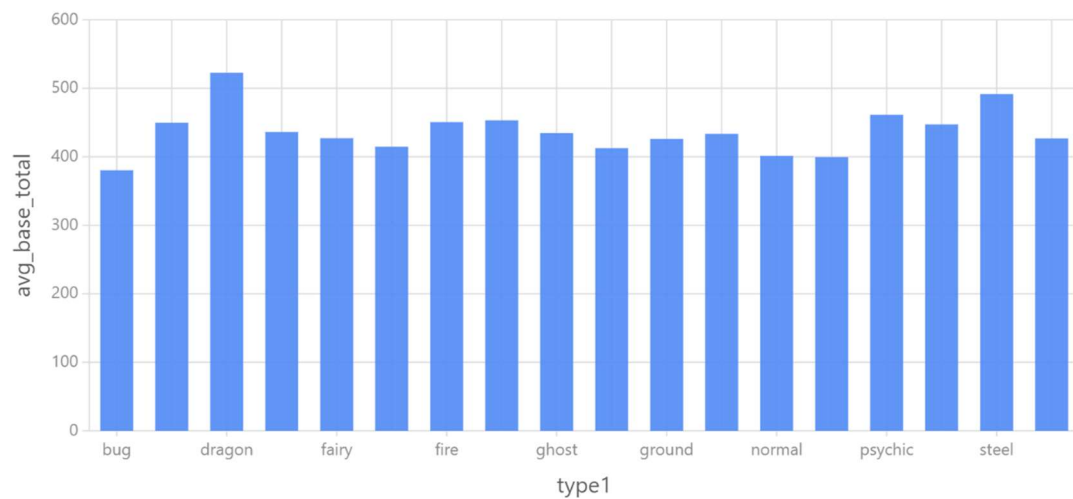
结论： 在对特殊技能抗性方面，最强的是 Artificial 类，或钢铁属性；最弱的是 Blade Quill 类，或岩石属性。在基础点数方面，最强的是 Alpha 类，或龙属性；最弱的是 Acorn 类，或虫属性。除此之外，对宝可梦的双属性对两指标的共作用分析详见下方表格。

分析：

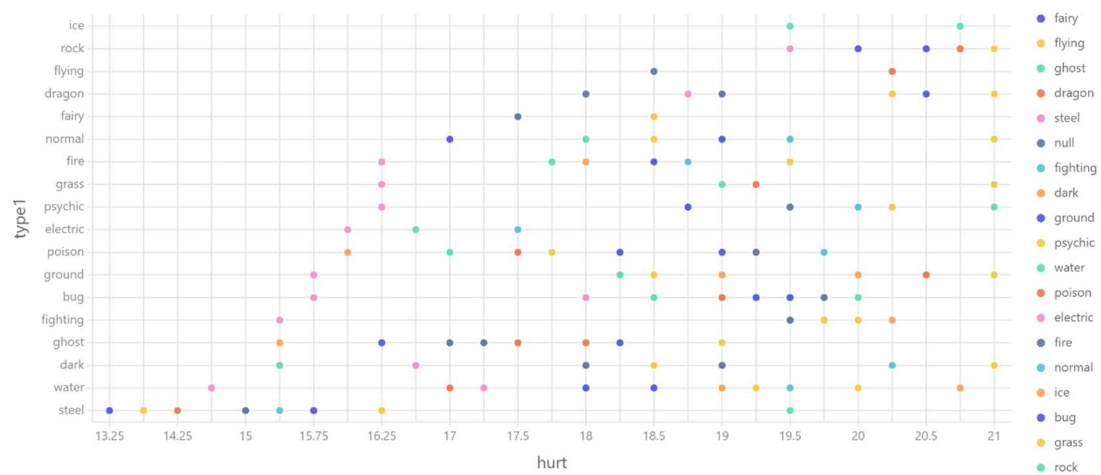
与上个问题相同，我们取“基础属性”和“抗性”作为宝可梦的能力评判标准。需要确定“类型”的指标。根据数据，每只宝可梦有一个“种类 (classification)” 标签和两个“属性 (type1, type2)” 标签。

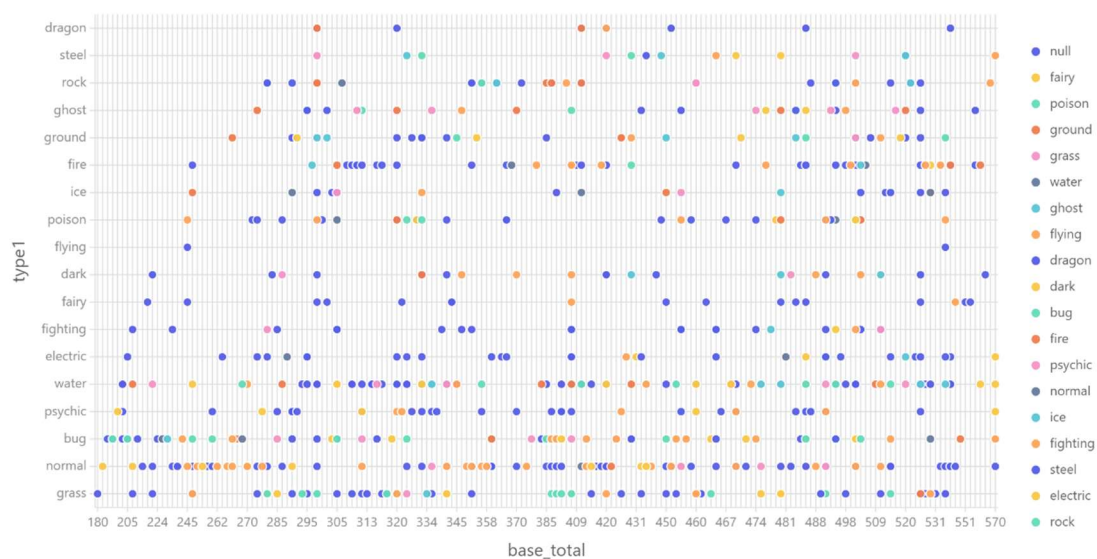
对每个种类计算伤害和基础属性平均值，作可视化图表得：





读表可知，龙属性宝可梦的基础点数最高，为 522.78；虫属性宝可梦基础点数最低，为 380.43。





除此之外，平台还可以对两个属性同时生成可视化图表。

读表可知，第一属性为钢铁，第二属性为精灵的宝可梦伤害最低（抗性最高），为13.25；伤害最高（抗性最低）的宝可梦有多种，分别为岩石1草2，龙1物理2，普通1草2，草1无2，物理1幽灵2，地面1物理2，黑暗1物理2。

基础点数最高的宝可梦有多种，分别为钢铁1飞行2，电1妖精2，水1妖精2，物理1妖精2，虫1格斗2，普通1无2，基础属性值均为570；最低的为草1无2，为180。

3. 身体特征与力量状态（进攻，防守，速度等）之间是否有相关性？

结论：身高和体重与部分力量状态有中度相关关系。

分析：

每个格子试相关性分析：（将特殊攻击、特殊防御与攻击、防御分别加和合并）

	攻击	防御	速度	体力
身高	0.64	0.54	0.31	0.61

体重	0.54	0.54	0.17	0.57
雄性占比	1.06	1.16	1.1	1.14
基础心情	1.19	1.16	1.1	1.07

可能由于平台计算失误，表格中有一半的相关系数大于1，我们将这部分数据视为无效数据。但表格中仍有一半数据具有参考价值。

对相关系数 r 的意义进行解读，如果两个参数 x 与 y ：

(1) 正相关，则 $r > 0$ 。且一般地，如果 $0.5 \leq |r| < 0.8$ 为 中度相关， $0.3 \leq |r| < 0.5$ 为 低度相关；

(2) 负相关，则 $r < 0$ ；

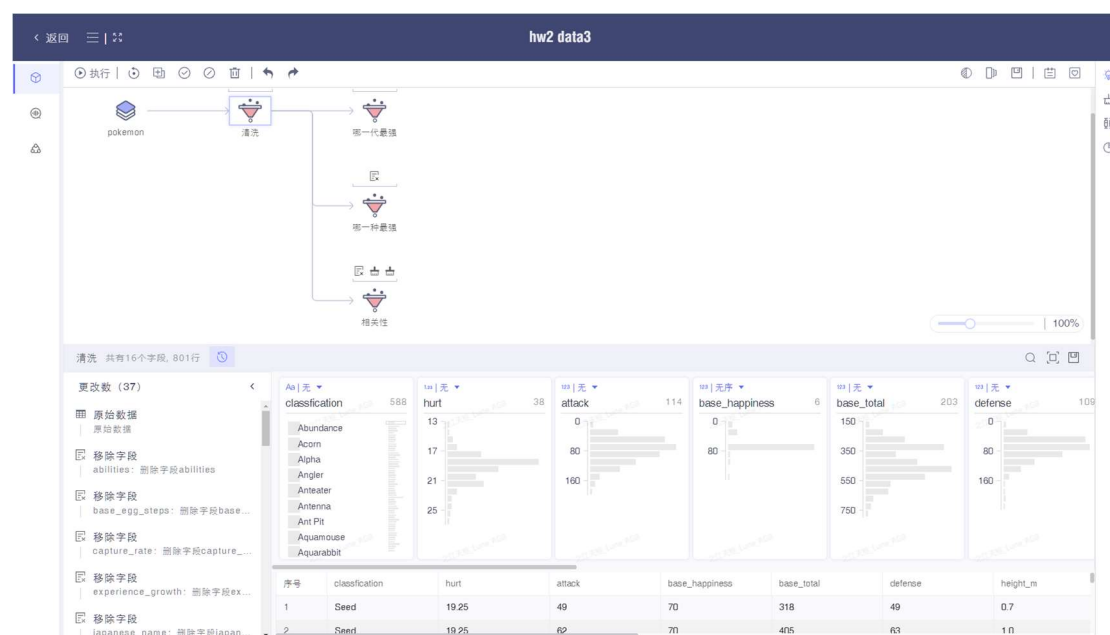
(3) 无线性关系，则 $r = 0$ 。

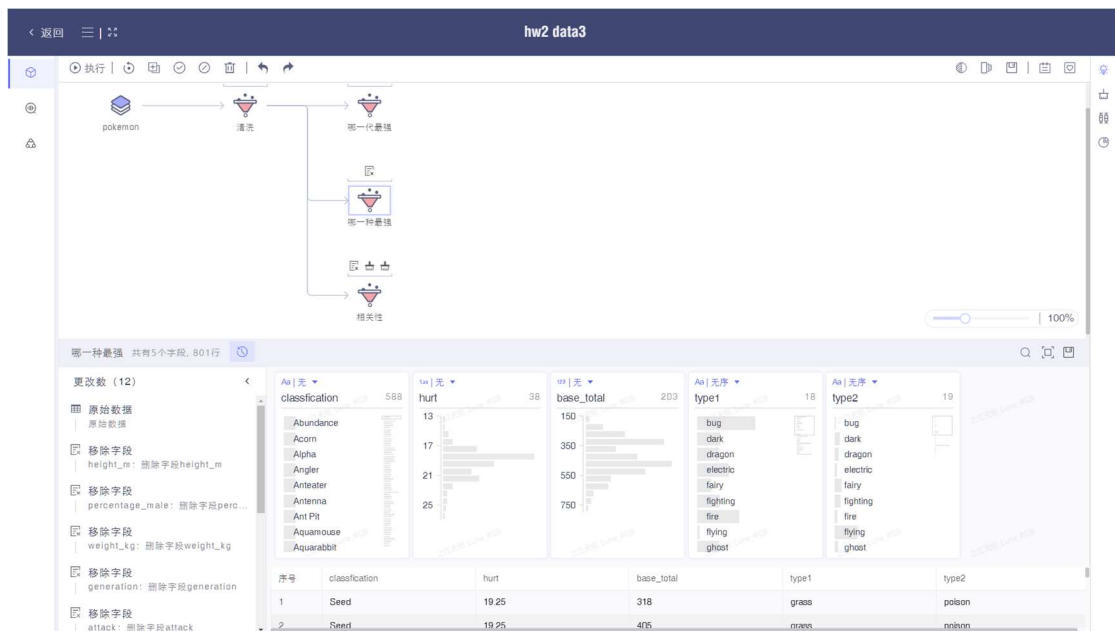
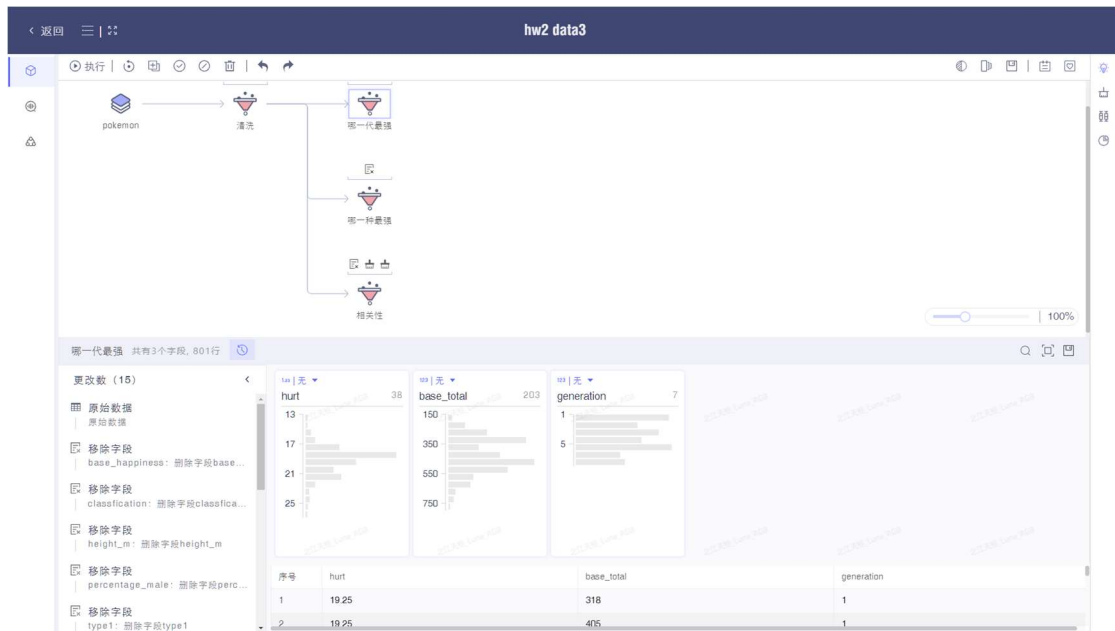
参照前两行，则身高与攻击、防御，体重与攻击、防御和体力均有中度相关关系。

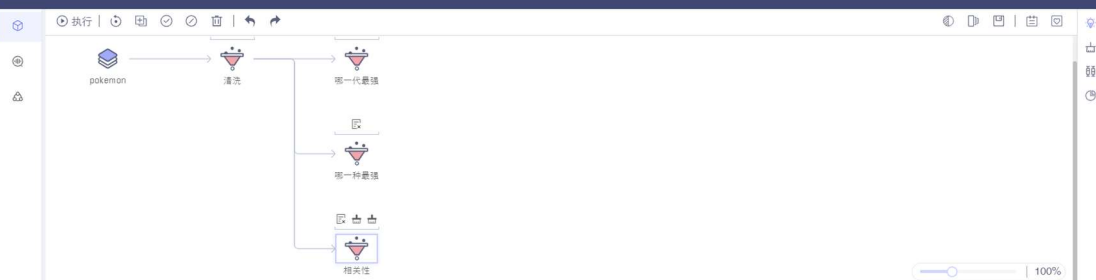
附录：

天枢平台工程文件链接及截图。

<https://nebula.zjvis.net/project/244/data>







相关性 共有8个字段, 801行

更改數 (31)

III 原始数据

◎ 雜誌訂閱

Ex 移除字段
classification: 删除字段classifica...

⑤ 移除字段

base_total: 删除字段base_total

⑤ 移除字段

type1: 删除子段type1

5x 移除字段

Ex	移陈字段
	generation



序号	base_happiness	height_m	hp	male_percentage	attack_total	defence_total	speed
1	70	0.7	45	0.8810000000000000...	114	114	45
2	70	1.0	60	0.8810000000000000...	142	143	60