
Exploring Cannabis Data from Leafly.com



A dank analysis by Kevin McMorro, Conner
Sexton

Cannabis Strains Dataset



```
> glimpse(df)
Observations: 2,350
Variables: 6
$ Strain      <fct> 100-0g, 98-White-Widow, 1024, 13-Dawgs, 24K-Gold, 3-Bears-0g, 3-Kings, 303-0g, 3D-Cbd, ...
$ Type        <fct> hybrid, hybrid, sativa, hybrid, hybrid, indica, hybrid, indica, sativa, indica, hybrid,...
$ Rating      <dbl> 4.0, 4.7, 4.4, 4.2, 4.6, 0.0, 4.4, 4.2, 4.6, 4.4, 4.7, 4.5, 4.4, 4.3, 4.3, 4.7, 5.0, 4...
$ Effects     <fct> "Creative,Energetic,Tingly,Euphoric,Relaxed", "Relaxed,Aroused,Creative,Happy,Energetic...
$ Flavor      <fct> "Earthy,Sweet,Citrus", "Flowery,Violet,Diesel", "Spicy/Herbal,Sage,Woody", "Apricot,Cit...
$ Description <fct> "$100 0G is a 50/50 hybrid strain that packs a strong punch. The name supposedly refers...
```

- Cannabis data on 2,350 unique strains
- Type: indica, sativa, or hybrid (indica & sativa)
- Rating: Leafly's 0-5 user-rating scale (avg. score)
- Effects: 13 unique effects
- Flavors: 49 unique flavors
- Description: background description

Data Management



Strain	Type	Rating	Effects	Flavor
1 100-Og	hybrid	4.0	Creative,Energetic,Tingly,Euphoric,Relaxed	Earthy,Sweet,Citrus
2 98-White-Widow	hybrid	4.7	Relaxed,Aroused,Creative,Happy,Energetic	Flowerly,Violet,Diesel
3 1024	sativa	4.4	Uplifted,Happy,Relaxed,Energetic,Creative	Spicy/Herbal,Sage,Woody
4 13-Dawgs	hybrid	4.2	Tingly,Creative,Hungry,Relaxed,Uplifted	Apricot,Citrus,Grapefruit
5 24K-Gold	hybrid	4.6	Happy,Relaxed,Euphoric,Uplifted,Talkative	Citrus,Earthy,Orange
6 3-Bears-Og	indica	0.0	None	None
7 3-Kings	hybrid	4.4	Relaxed,Euphoric,Happy,Uplifted,Hungry	Earthy,Sweet,Pungent
8 303-Og	indica	4.2	Relaxed,Happy,Euphoric,Uplifted,Giggly	Citrus,Pungent,Earthy
9 3D-Cbd	sativa	4.6	Uplifted,Focused,Happy,Talkative,Relaxed	Earthy,Woody,Flowery
10 3X-Crazy	indica	4.4	Relaxed,Tingly,Happy,Euphoric,Uplifted	Earthy,Grape,Sweet

	Creative	Energetic	Tingly	Euphoric	Relaxed	Aroused	Happy	Uplifted
100-Og	1	1	1	1	1	0	0	0
98-White-Widow	1	1	0	0	1	1	1	0
1024	1	1	0	0	1	0	1	1
13-Dawgs	1	0	1	0	1	0	0	1
24K-Gold	0	0	0	1	1	0	1	1
3-Bears-Og	0	0	0	0	0	0	0	0
3-Kings	0	0	0	1	1	0	1	1
303-Og	0	0	0	1	1	0	1	1
3D-Cbd	0	0	0	0	1	0	1	1
3X-Crazy	0	0	1	1	1	0	1	1
3Rd-Coast-Panama-Chunk	0	1	0	0	0	0	1	1
501St-Og	0	0	0	1	1	0	1	0
5Th-Element	0	0	0	1	1	1	0	1
707-Headband	1	0	0	1	0	0	1	1
8-Ball-Kush	1	0	0	1	1	0	1	1

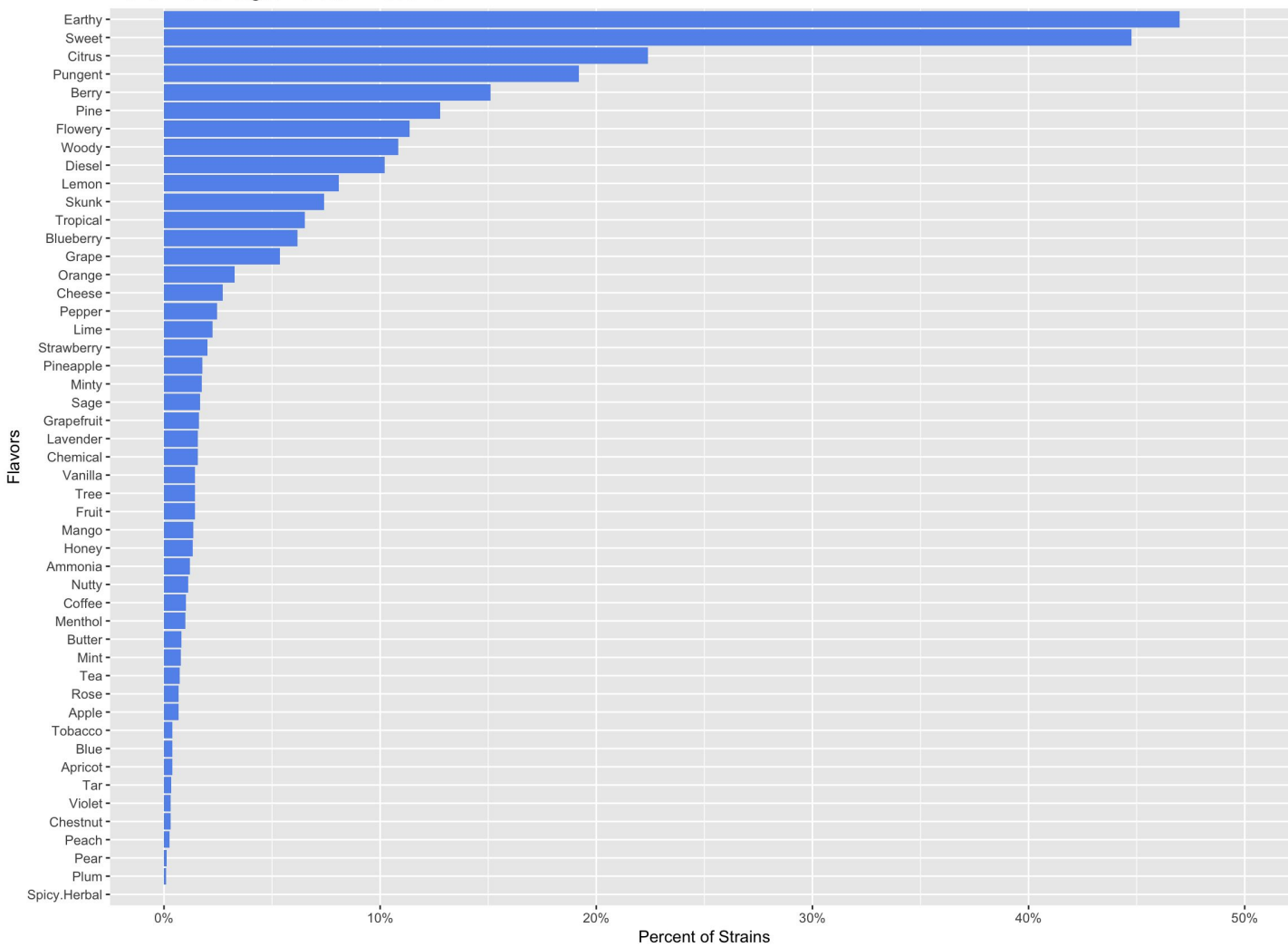
	Earthy	Sweet	Citrus	Flowery	Violet	Diesel	Spicy.Herbal	Sage	Woody
100-Og	1	1	1	0	0	0	0	0	0
98-White-Widow	0	0	0	1	1	1	0	0	0
1024	0	0	0	0	0	0	0	1	1
13-Dawgs	0	0	1	0	0	0	0	0	0
24K-Gold	1	0	1	0	0	0	0	0	0
3-Bears-Og	0	0	0	0	0	0	0	0	0
3-Kings	1	1	0	0	0	0	0	0	0
303-Og	1	0	1	0	0	0	0	0	0
3D-Cbd	1	0	0	1	0	0	0	0	1
3X-Crazy	1	1	0	0	0	0	0	0	0
3Rd-Coast-Panama-Chunk	0	0	0	0	0	1	0	0	0
501St-Og	1	0	0	0	0	0	0	0	0
5Th-Element	1	0	0	0	0	0	0	0	1
707-Headband	1	0	0	0	0	0	0	0	1
8-Ball-Kush	1	0	1	0	0	0	0	0	0

Our Guiding Questions

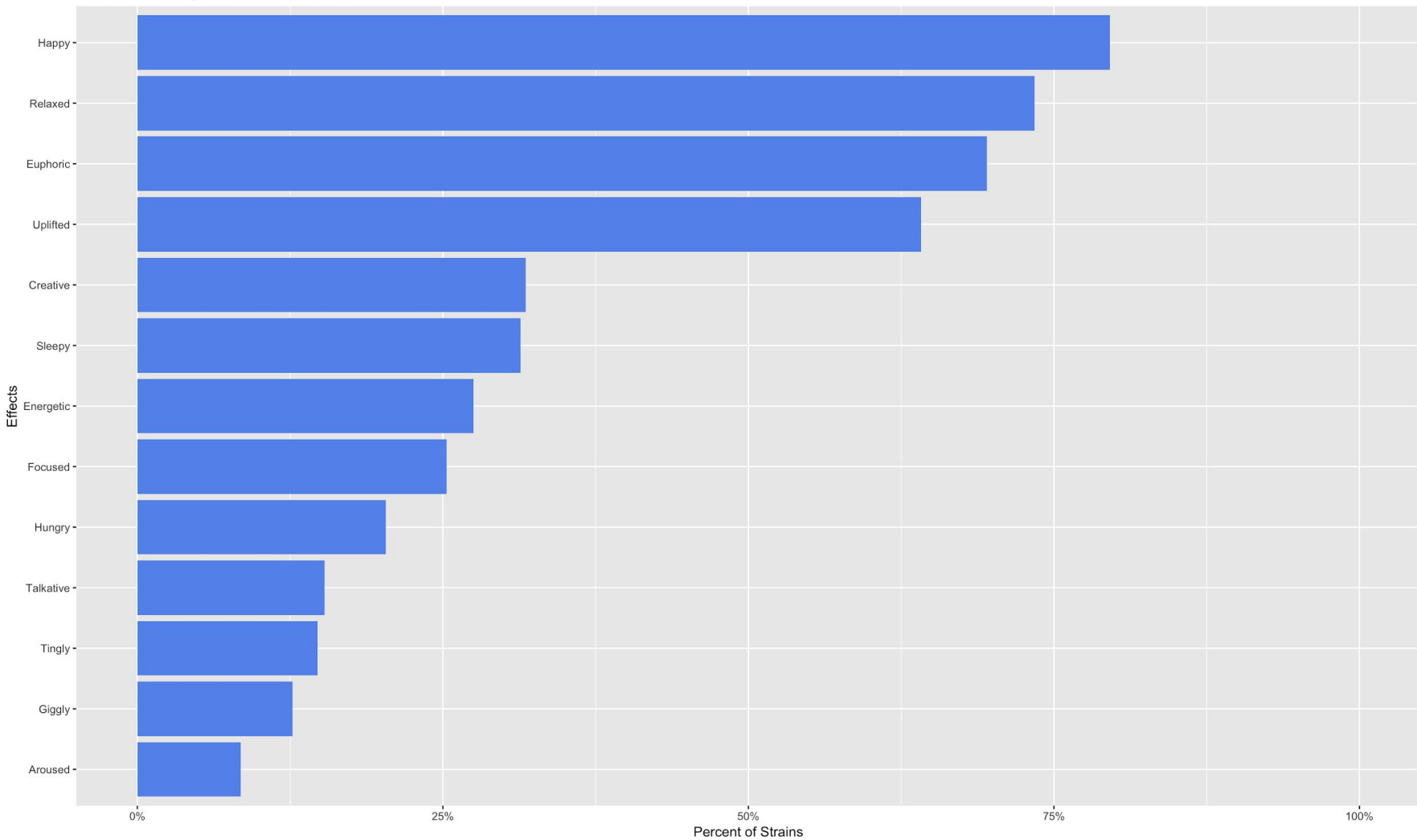


- How do flavor and effect influence the user rating of a cannabis strain?
- How do strain types (indica, hybrid, sativa) differ in terms of flavor and effect compositions?
- Are there underlying factors within the data? If so, what are they?
- What effects/flavors are associated with highly rated strains?

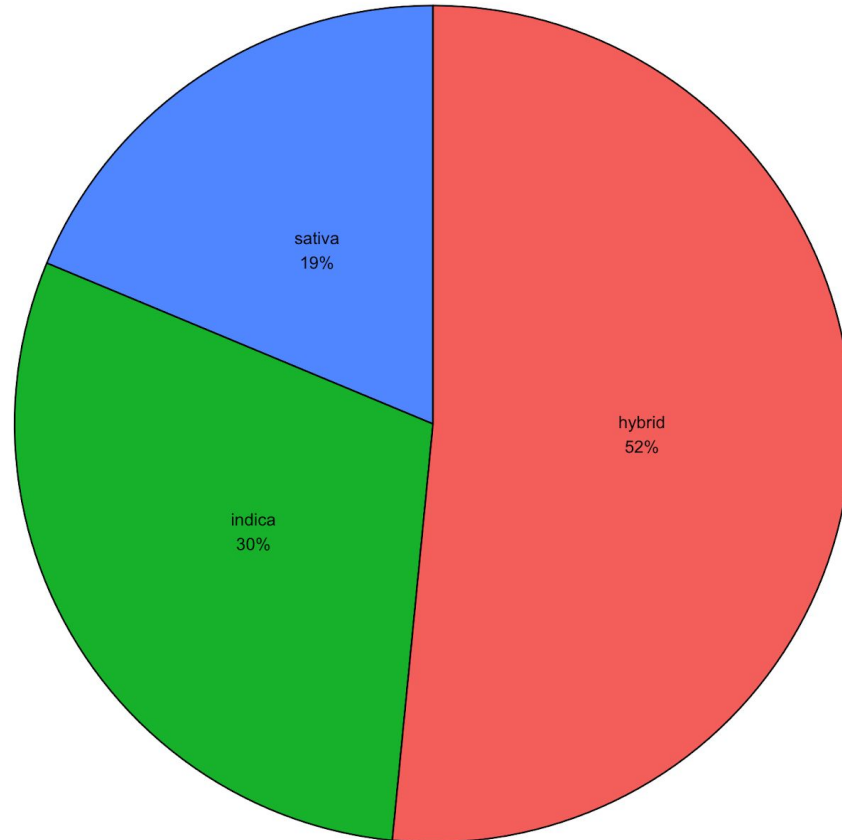
Flavor Percentages Across All Strains



Effect Percentages Across All Strains



Strain Type Percentages



Factor Analysis:

- What characteristics of flavor and effect can we represent through factors?

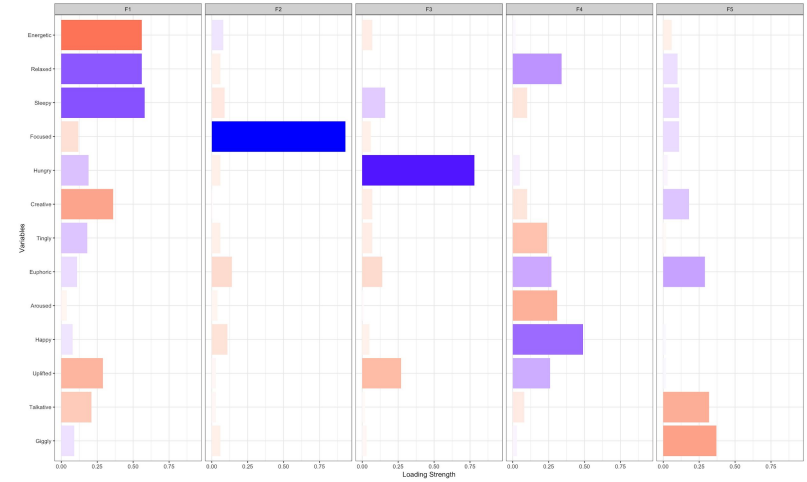


Factor Analysis (Strain Effects)

Total variance explained: 0.31

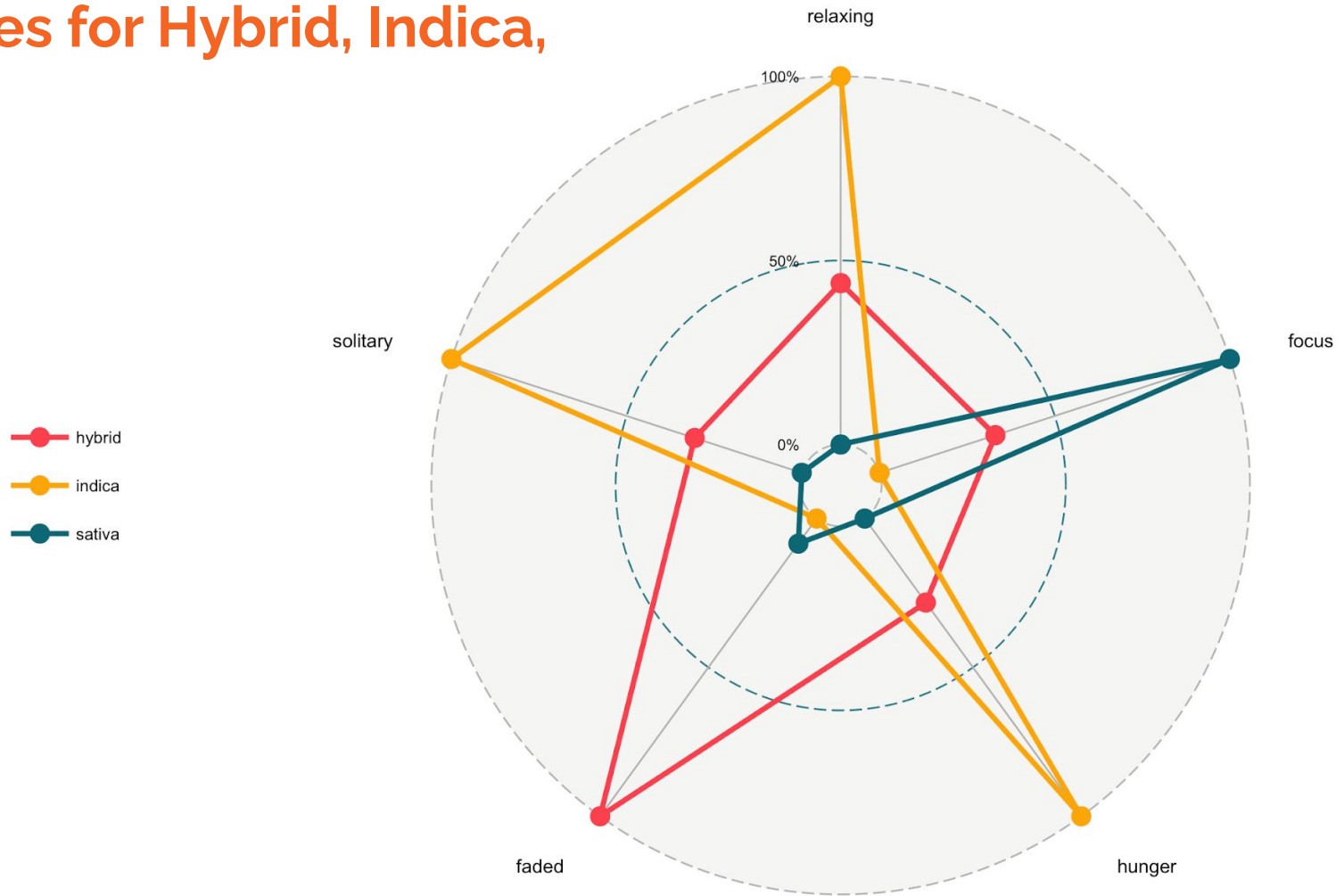
Lowest Communality: Tingly/Aroused (0.10)

Highest Communality: Focused (0.89)



	F1. Relaxing	F2. Focus	F3. Hunger	F4. Faded	F5. Solitary
High Correlations	Relaxed, Sleepy	Focused	Hungry	Happy, Uplifted, Relaxed, Euphoric	Euphoric
Low Correlations	Energetic, Creative, Uplifted	-	Uplifted	Tingly, Aroused	Talkative, Giggly

Effect Profiles for Hybrid, Indica, Sativa



Factor Analysis (Strain Flavors)

Total variance explained: 0.12

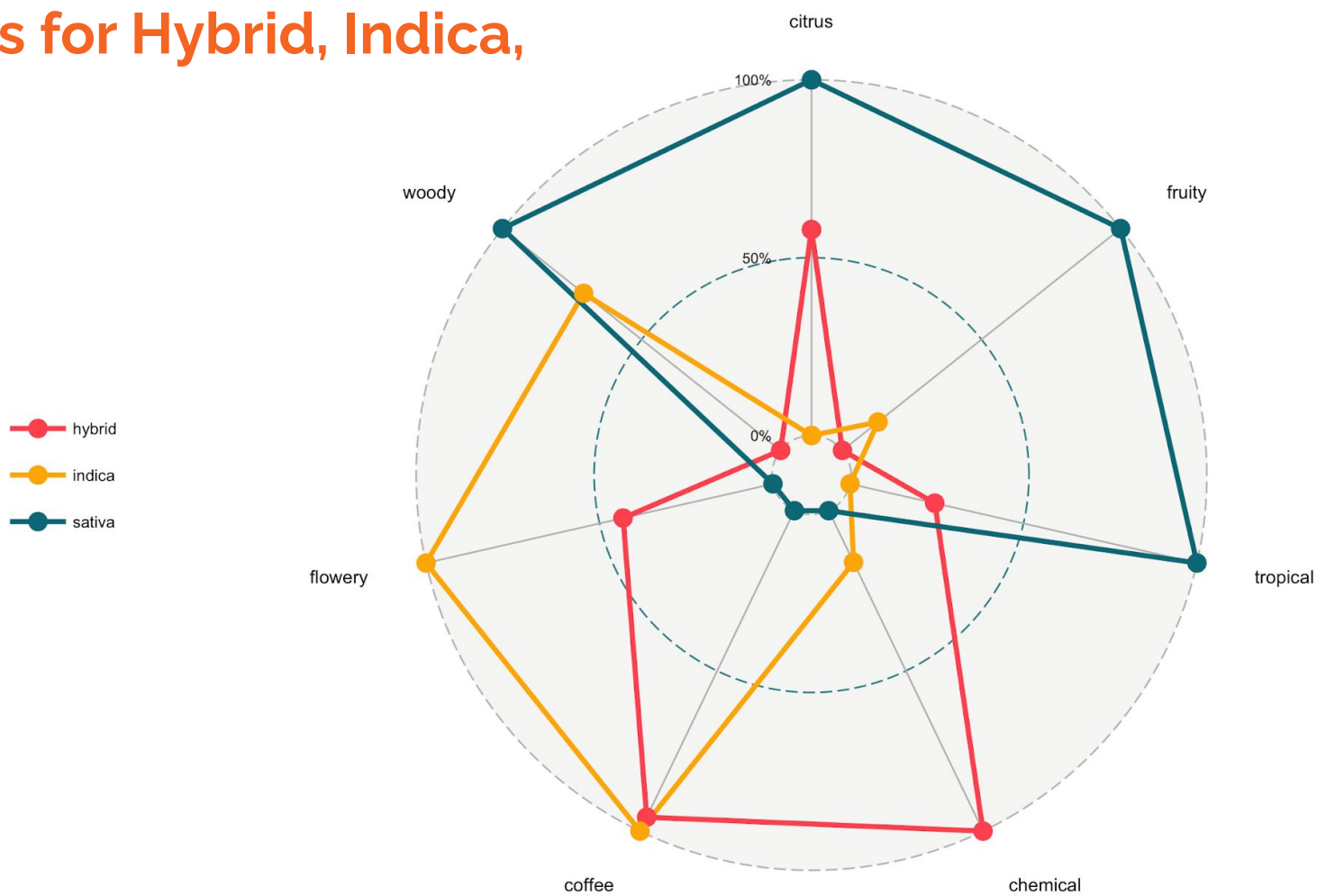
Lowest Communality: Fruit (0.10)

Highest Communality: Citrus (0.89)



	1. Citrus	2. Fruity	3. Cheese	4. Tropical	5. Chemical	6. Coffee	7. Flowery	8. Woody
High Correlations	Citrus, Orange, Lemon, Lime	Berry, Sweet, Blueberry	Cheese, Blue	Tropical, Pineapple	Ammonia, Diesel, Menthol	Vanilla, Coffee	Lavender, Rose, Flowery	Woody, Pine
Low Correlations	Cheese, Earthy	Earthy, Pungent	Earthy	Citrus, Berry	Earthy, Sweet	Tropical	Vanilla	Pungent, Skunk

Flavor Profiles for Hybrid, Indica, Sativa



Regression Analysis on Effect and Flavor Factors.

Call:

```
lm(formula = rating ~ factor(type) + relaxing + focus + hunger + faded + solitary + citrus + fruity + cheese + tropical + chemical + coffee + flowery + woody, data = complete)
```

Residuals:

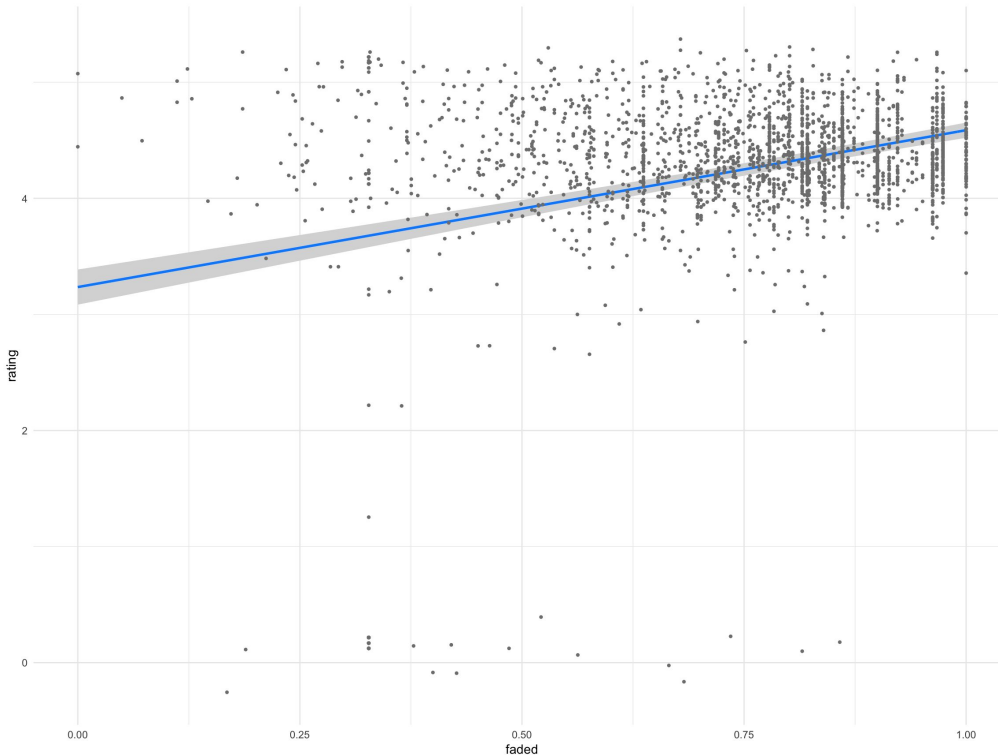
Min	1Q	Median	3Q	Max
-4.3218	-0.2067	0.0367	0.3549	1.8387

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.89539	0.19305	14.998	< 2e-16	***
factor(type)indica	0.09398	0.04165	2.256	0.02413	*
factor(type)sativa	0.04870	0.04603	1.058	0.29014	
relaxing	0.09058	0.08113	1.116	0.26432	
focus	0.22697	0.04875	4.656	3.40e-06	***
hunger	-0.01702	0.05519	-0.308	0.75787	
faded	1.35138	0.09143	14.781	< 2e-16	***
solitary	0.38613	0.09158	4.216	2.58e-05	***
citrus	0.18119	0.08008	2.263	0.02375	*
fruity	0.19144	0.08070	2.372	0.01775	*
cheese	0.02368	0.17560	0.135	0.89275	
tropical	-0.29330	0.11113	-2.639	0.00837	**
chemical	-0.31528	0.12588	-2.505	0.01233	*
coffee	-0.19613	0.17063	-1.149	0.25047	
flowery	0.27191	0.23126	1.176	0.23979	
woody	-0.02868	0.13732	-0.209	0.83461	

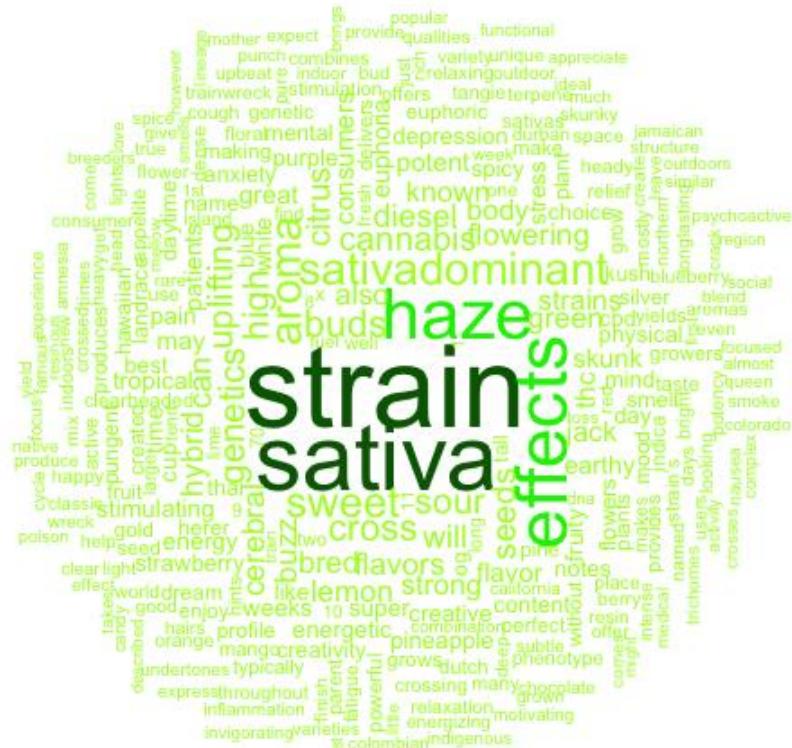
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.775 on 2334 degrees of freedom
Multiple R-squared: 0.1469, Adjusted R-squared: 0.1414
F-statistic: 26.8 on 15 and 2334 DF, p-value: < 2.2e-16



Strain Descriptions Summarized

(3)



>sample(df\$Description, 1)

[1] "Although it is a hybrid, the AK-48 high leans more towards typical sativa effects. AK-48's parents are Ice and Jock Horror. Its Ice mother has ancestors of Afghani, Skunk, Northern Lights and Shiva stock."

Limitations and Further Analysis:

- Researchers have argued that strain-type (Sativa/Indica/Hybrid) is not an effective means of explaining a strain's effects. Instead, it would be interesting to examine the chemical breakdown of a strain's *cannabinoids* (THC v CBD)
- New strains are constantly being added to the Leafly website, so analyzing strains in regards to user rating can be difficult when new strains tend to have <10 ratings
- Leafly allows users to leave lengthy reviews of strains. Finding a way to scrape the website for these reviews could prove interesting in future analysis.