

Exponential Convolution

By: Trinity Love



In High Regard



Gryphon Chinese Fire



Rose F. Kennedy



So I may
have a
slight
problem



Some of my seedlings



(Wild Horses x UNK)



(Unbridled x Kindly Light)?



(Sona's Song x Indian Giver)





(Indian Giver x Kindly Light)

Unknown
parentage
possibly with
Sona's Song and
or Final Touch



NOW FOR THE FUN STUFF

HOW TO STEAL A DATABASE



Web Scraping



Copy

- | Python Script
 - Selenium
 - BeautifulSoup

.CSV copy of the database

NOW WHAT?

...See what we got I guess?

99,243

Number of scraped varieties ±100 as of Dec 2023

That's A LOT of daylilies but what did we miss..?

(Web Scraping isn't totally consistent)

- Any variety that starts with a number 0-9 as the first character of their name
- Inconsistent scraping missed about 20ish entries information
- Withdrawn varieties

Daylily Database Search

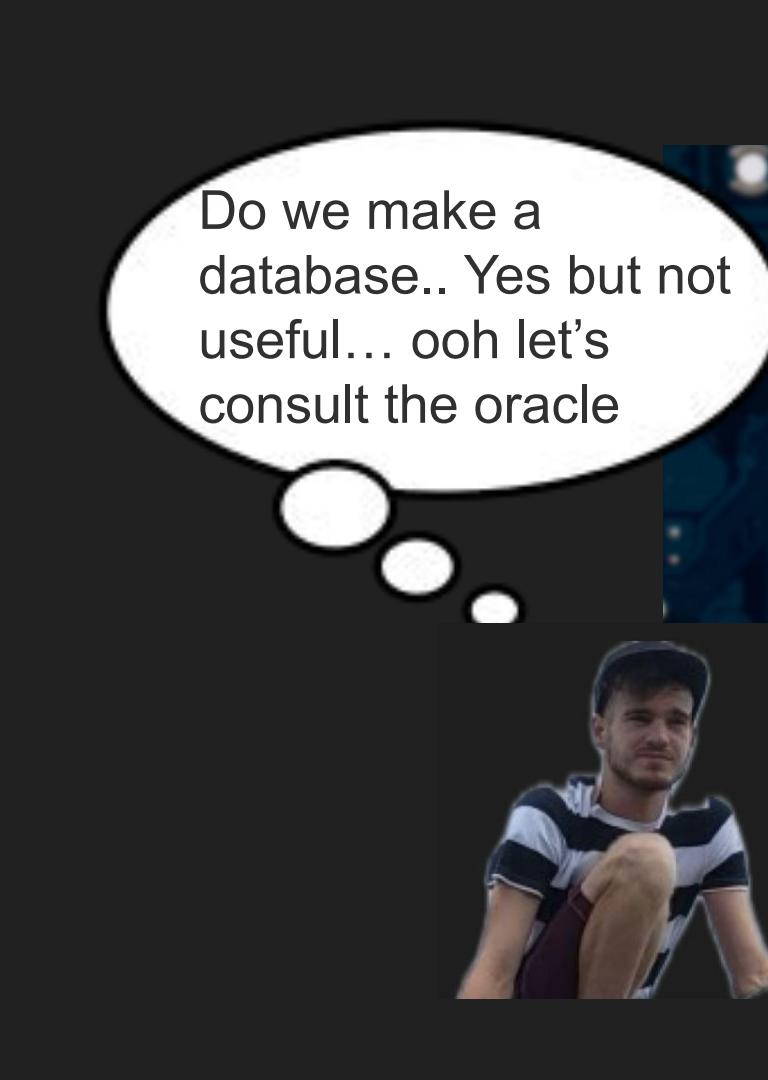
Search Results Page 1 of 1

1 daylilies found.
(showing 1 to 1)

Downton Abbey (WITHDRAWN: see Margaret Gibson, 0000)
height 0 inches (0.0 cm), , See MARGARET GIBSON.

[Print Results](#)

1 daylilies found. Page 1 of 1
(showing 1 to 1)

A composite image featuring a man in a striped shirt on the left and a hand holding a smartphone on the right. A thought bubble originates from the man's head, containing the text.

Do we make a
database.. Yes but not
useful... ooh let's
consult the oracle

OpenAI



DATA ANALYTICS

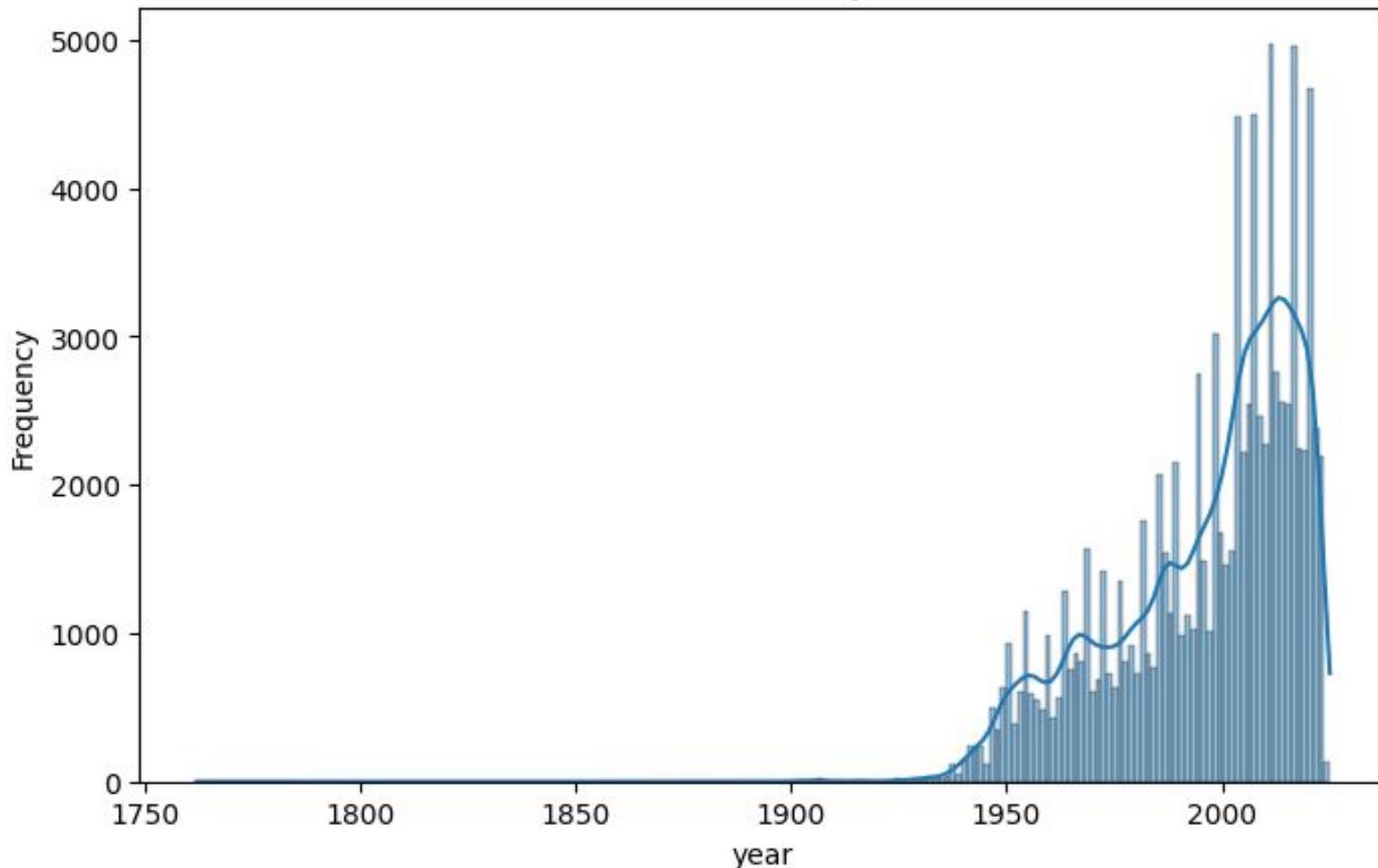
Using Google Co-Lab and PyCharm

(Python, Pandas, NumPy, Matplotlib.pyplot, and Seaborn)

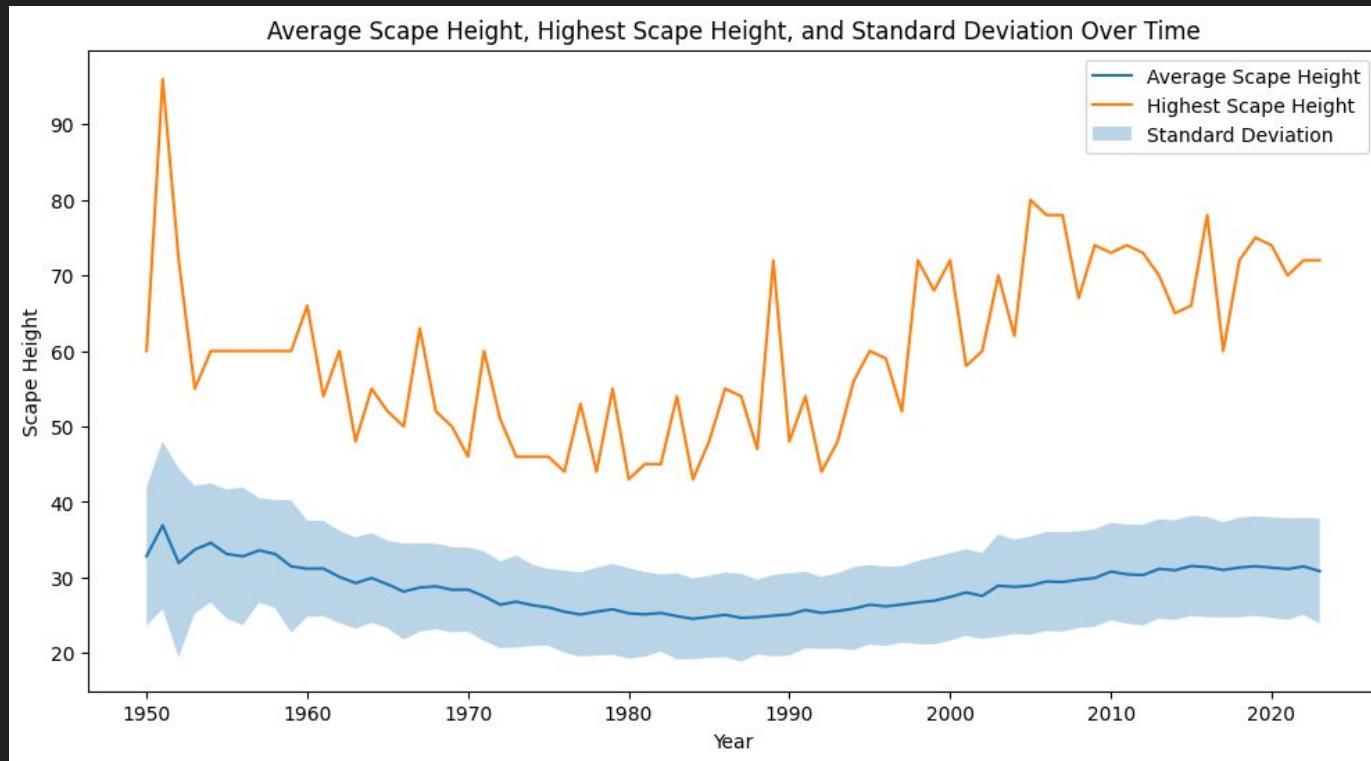
	variety	hybridizer	bloom_season	rebloom		form	ploidy
count	99242	99242	98478	44910		12500	97487
unique	99242	5277	7	1		432	2
top	A 2 B'	Wild	Midseason	Rebloom	Unusual Form Crispate	Tetraploid	
freq	1	1835	42543	44910		2765	50696

	leaf_type	bloom_habit	fragrance	seedling_number	parentage
count	96944	93367	32445	85109	77496
unique	3	5	2	77704	55932
top	Dormant	Diurnal	Fragrant	1 (unknown × unknown)	
freq	43554	70989	28847	227	4623

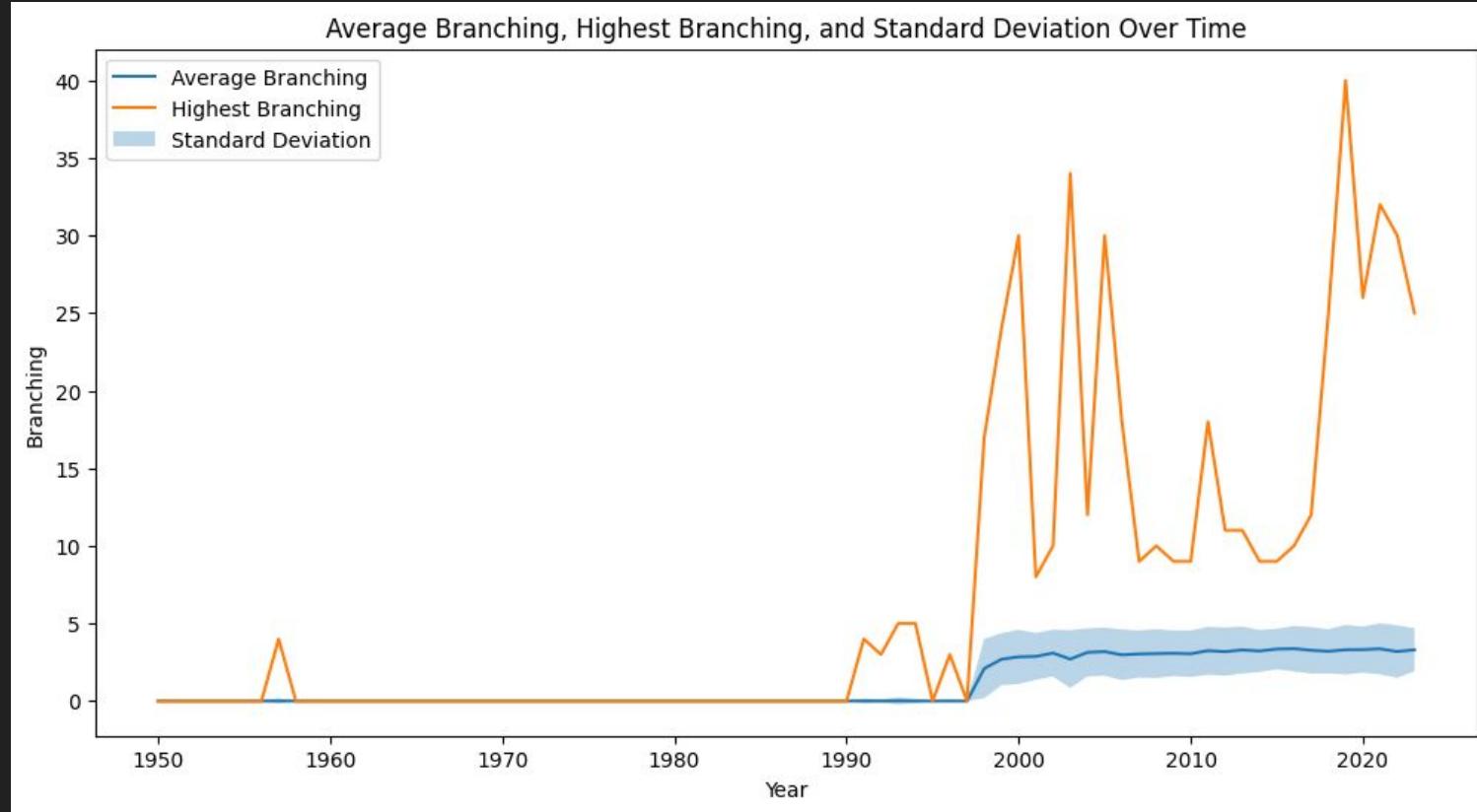
Distribution of year



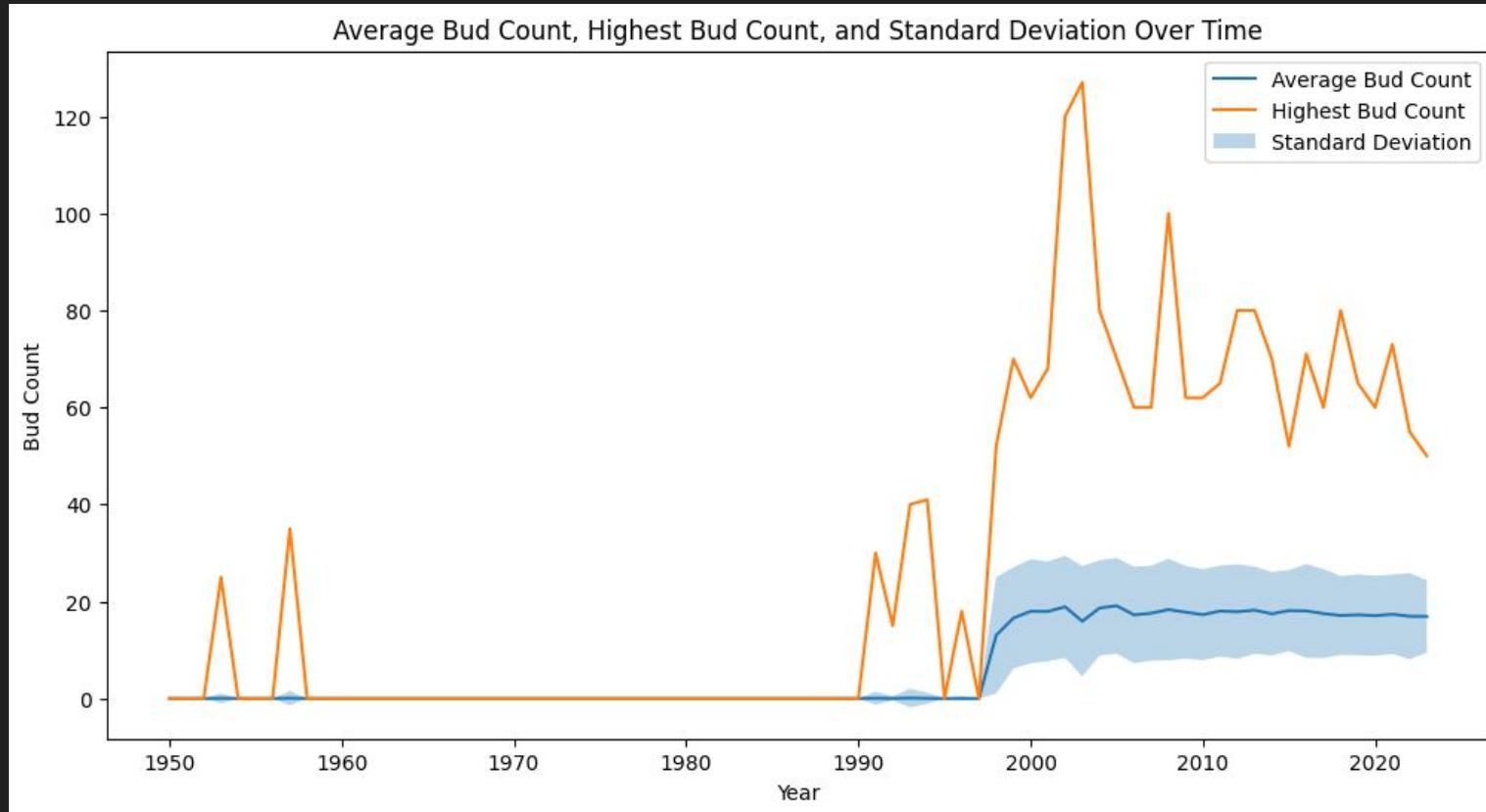
Scape Height - Mean: 29.114729650752704, Std Dev: 7.041889761032968,
Min: 0.0, Max: 96.0



Branching - Mean: 1.8208419822252675, Std Dev: 1.9523721314368587,
Min: 0.0, Max: 40.0

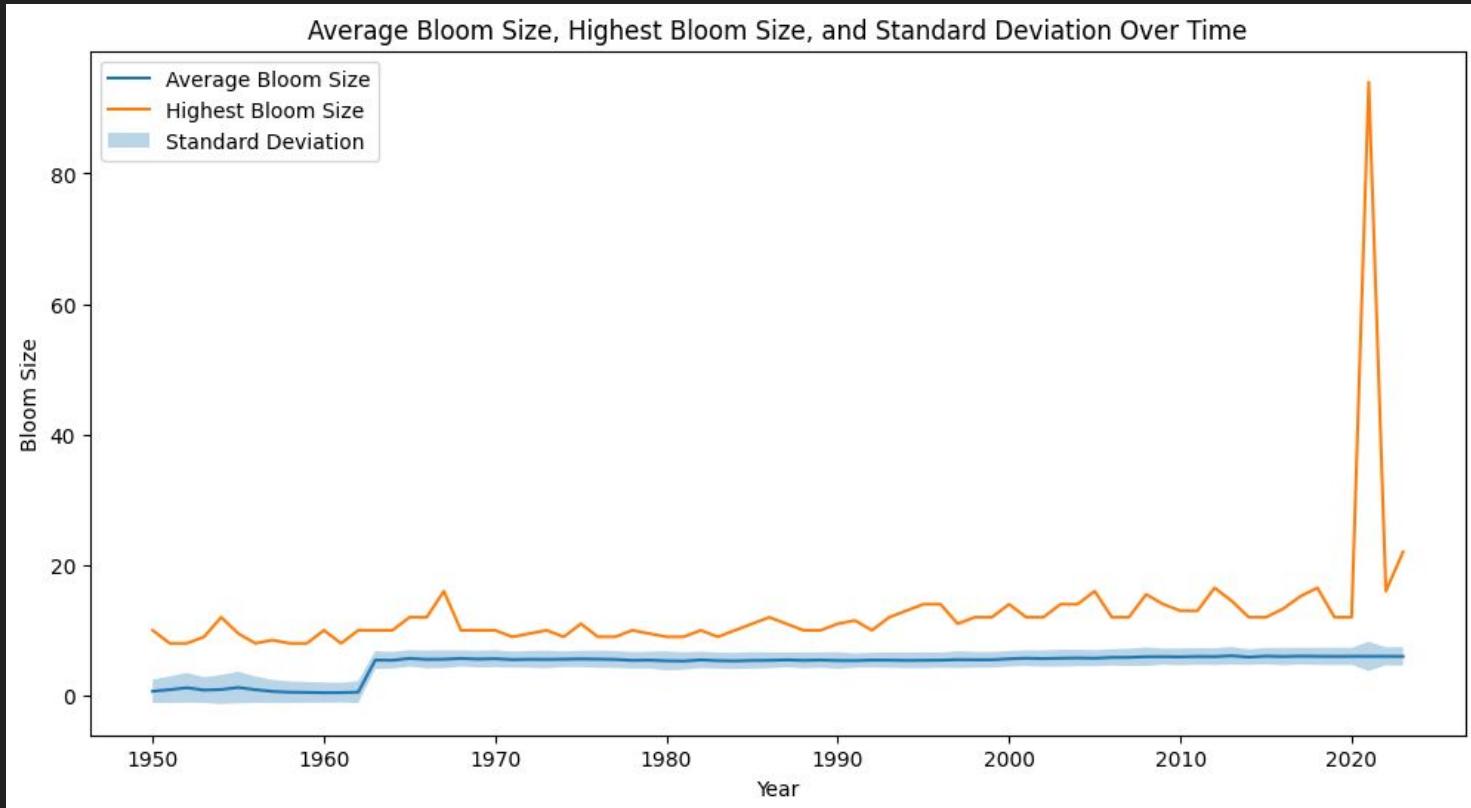


Bud Count - Mean: 10.23015457165313, Std Dev: 11.299265529748515,
Min: 0.0, Max: 127.0

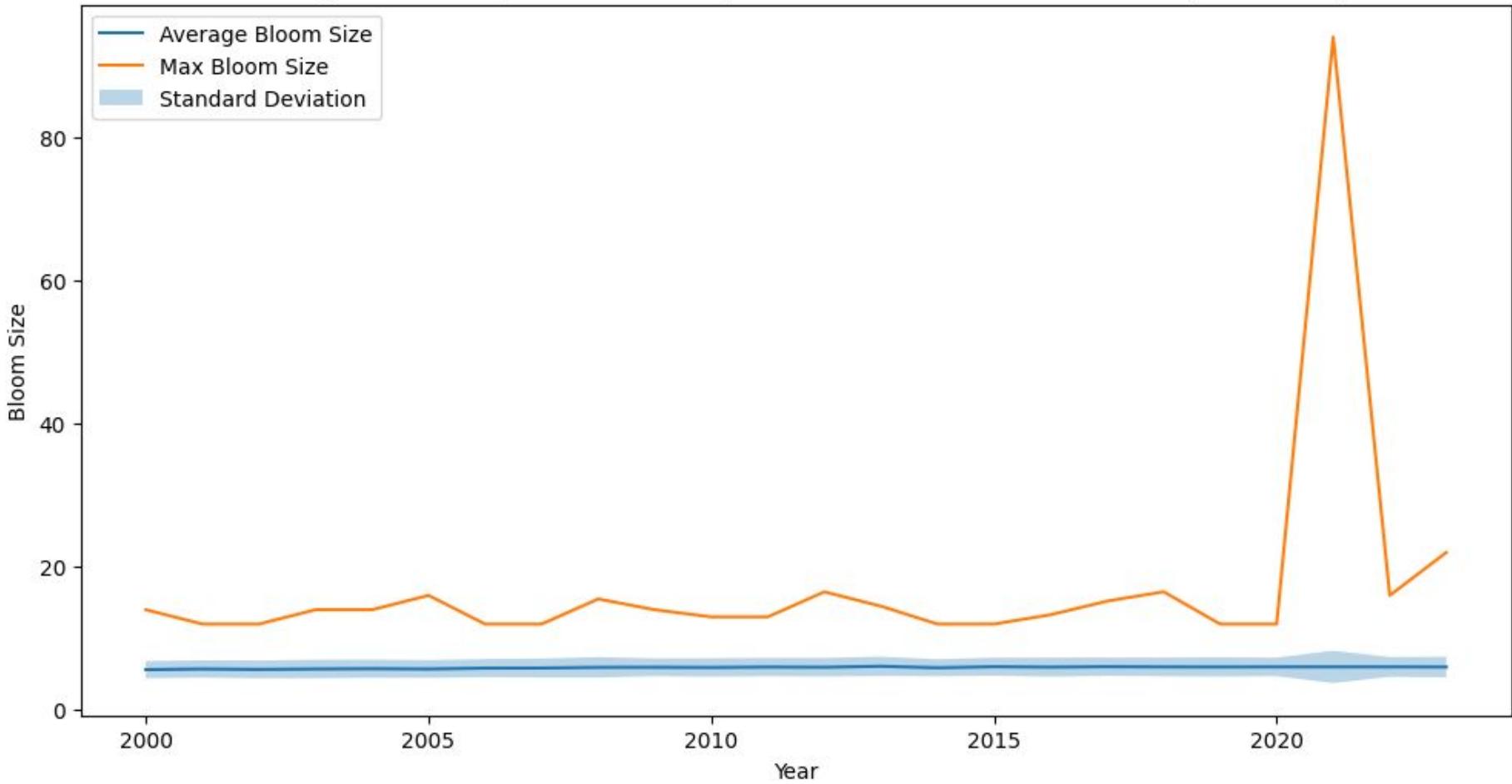


Bloom Size - Mean: 5.268068257391024, Std Dev: 2.0248868714046533,
Min: 0.0, Max: 94.0

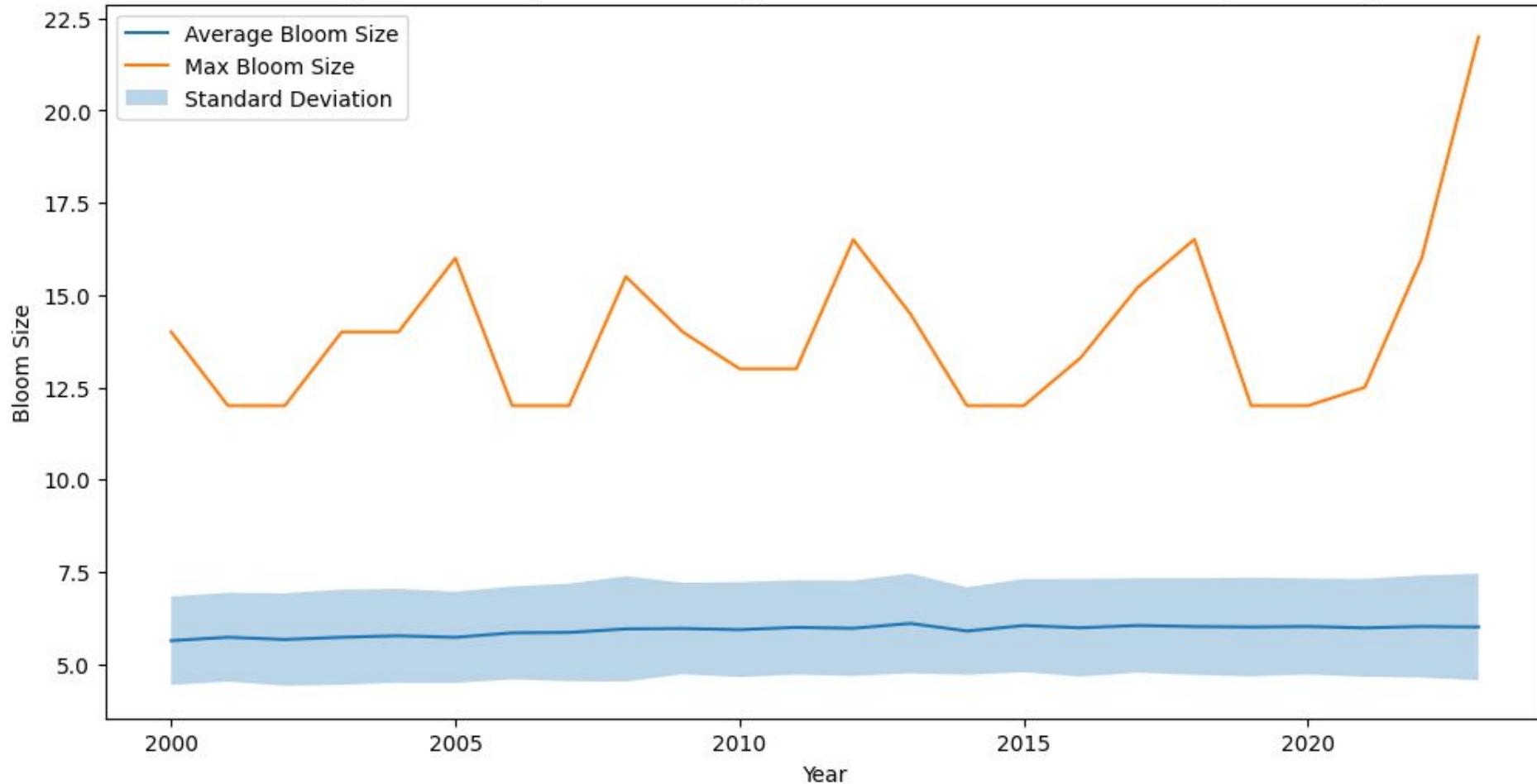
What's
wrong
here??



Average Bloom Size, Max Bloom Size, and Standard Deviation Over Time (2000-2023)



Average Bloom Size, Max Bloom Size, and Standard Deviation Over Time (2000-2023)

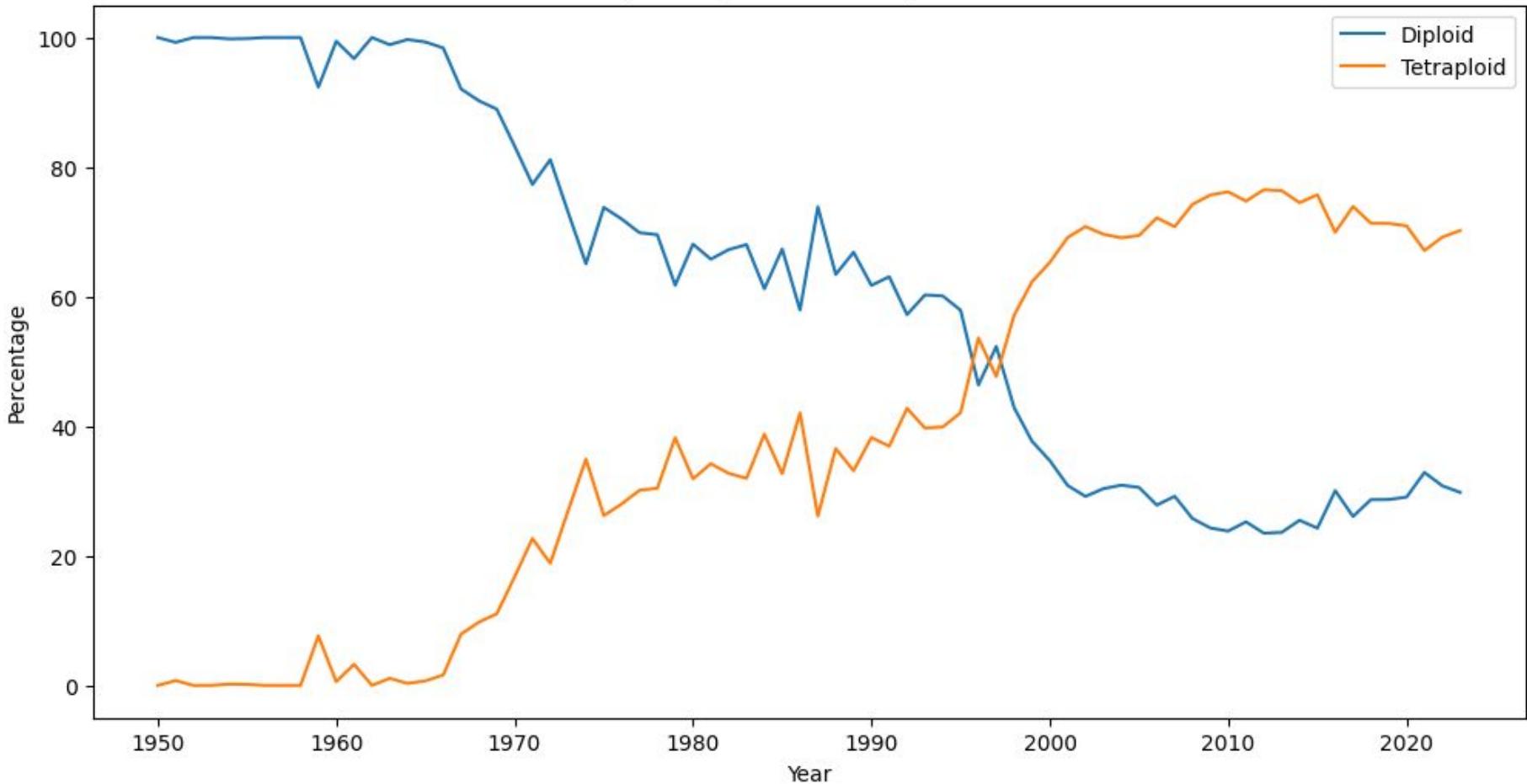


Percentage of each ploidy:

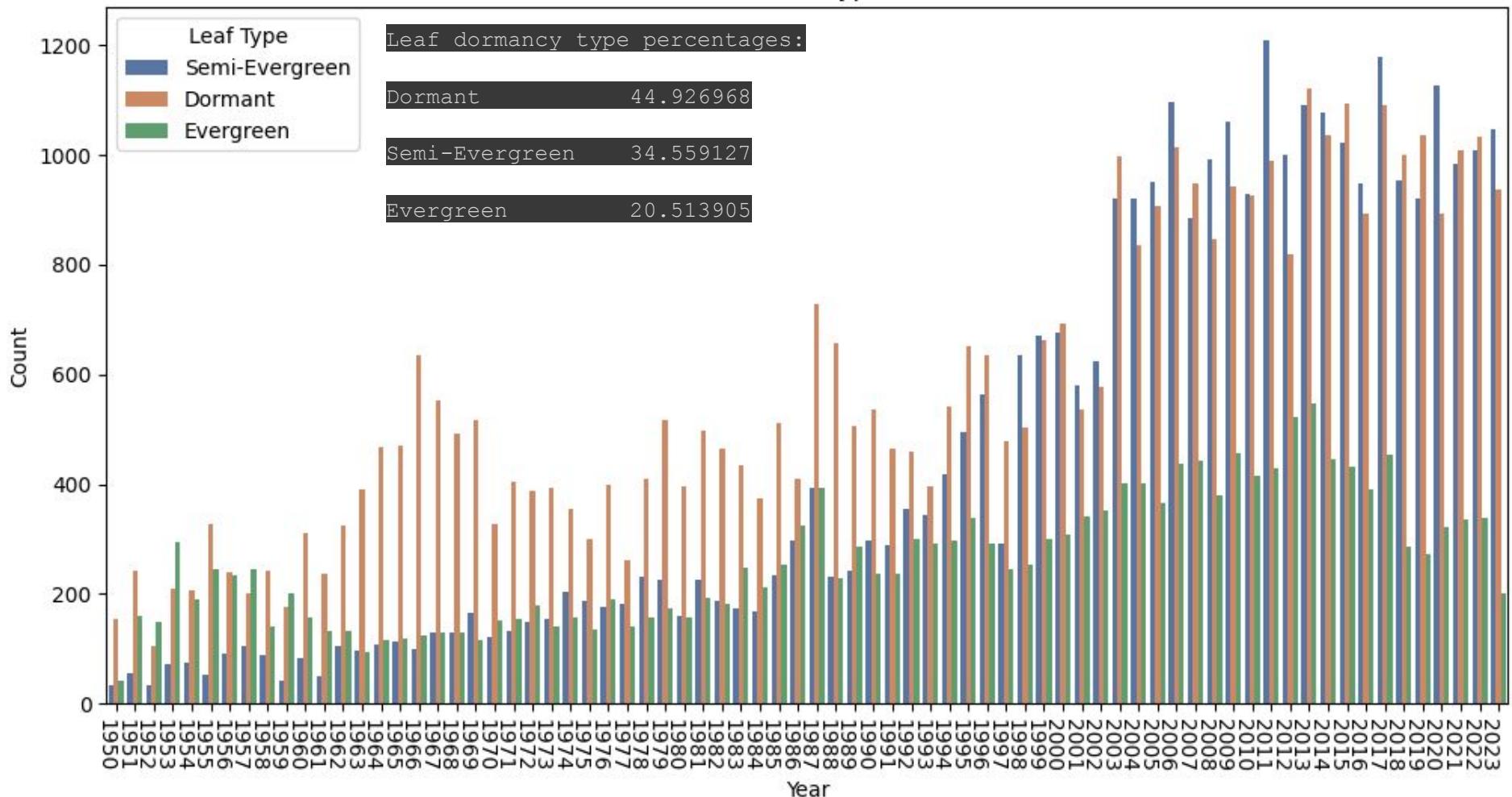
Tetraploid 52.002831

Diploid 47.997169

Percentage of Diploid and Tetraploid Over Time



Distribution of Leaf Types Over Time

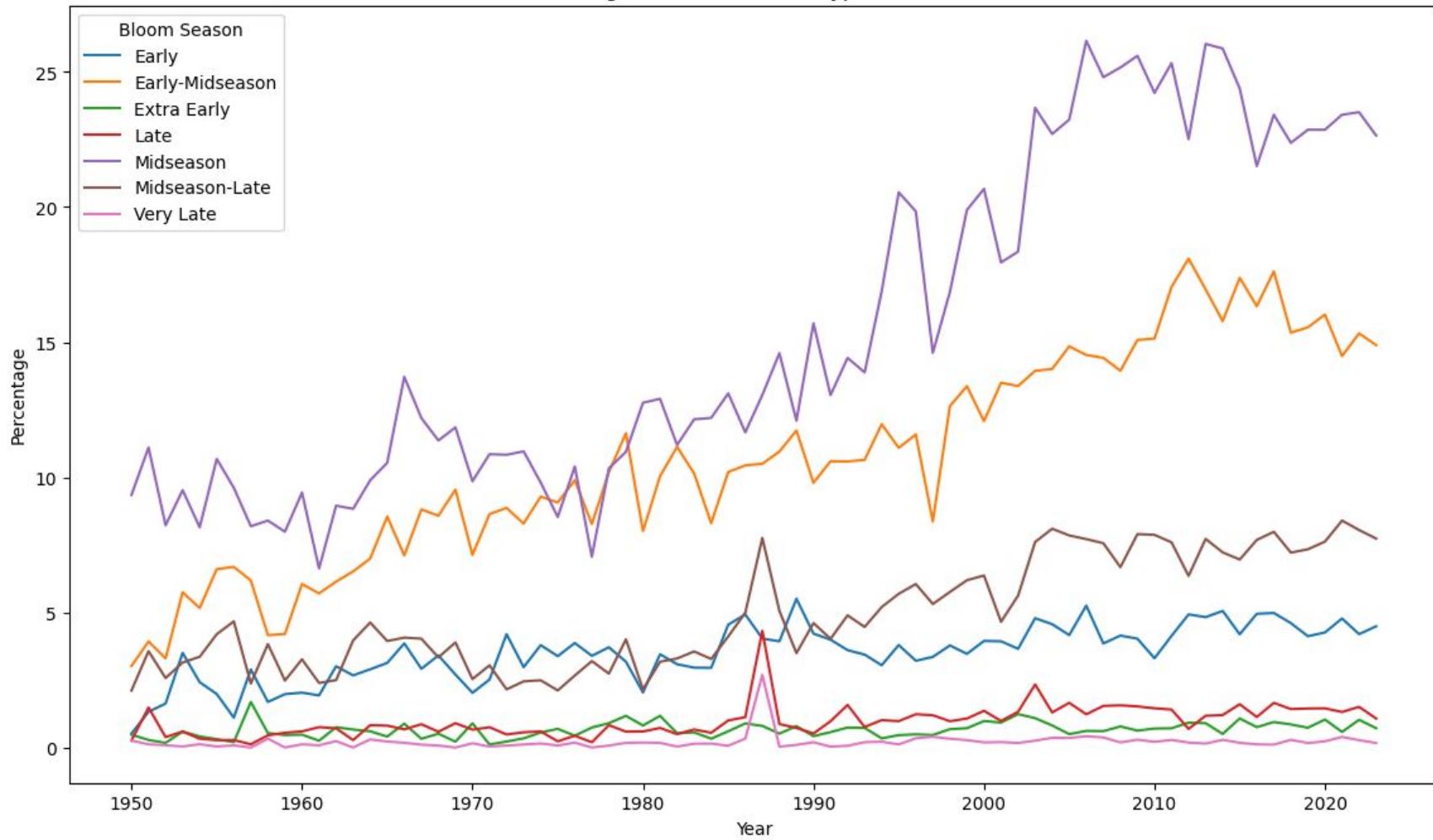


Bloom season percentages:

Midseason	43.200512
Early-Midseason	29.017649
Midseason-Late	13.604054
Early	9.233534
Late	2.671663
Extra Early	1.701903
Very Late	0.570686



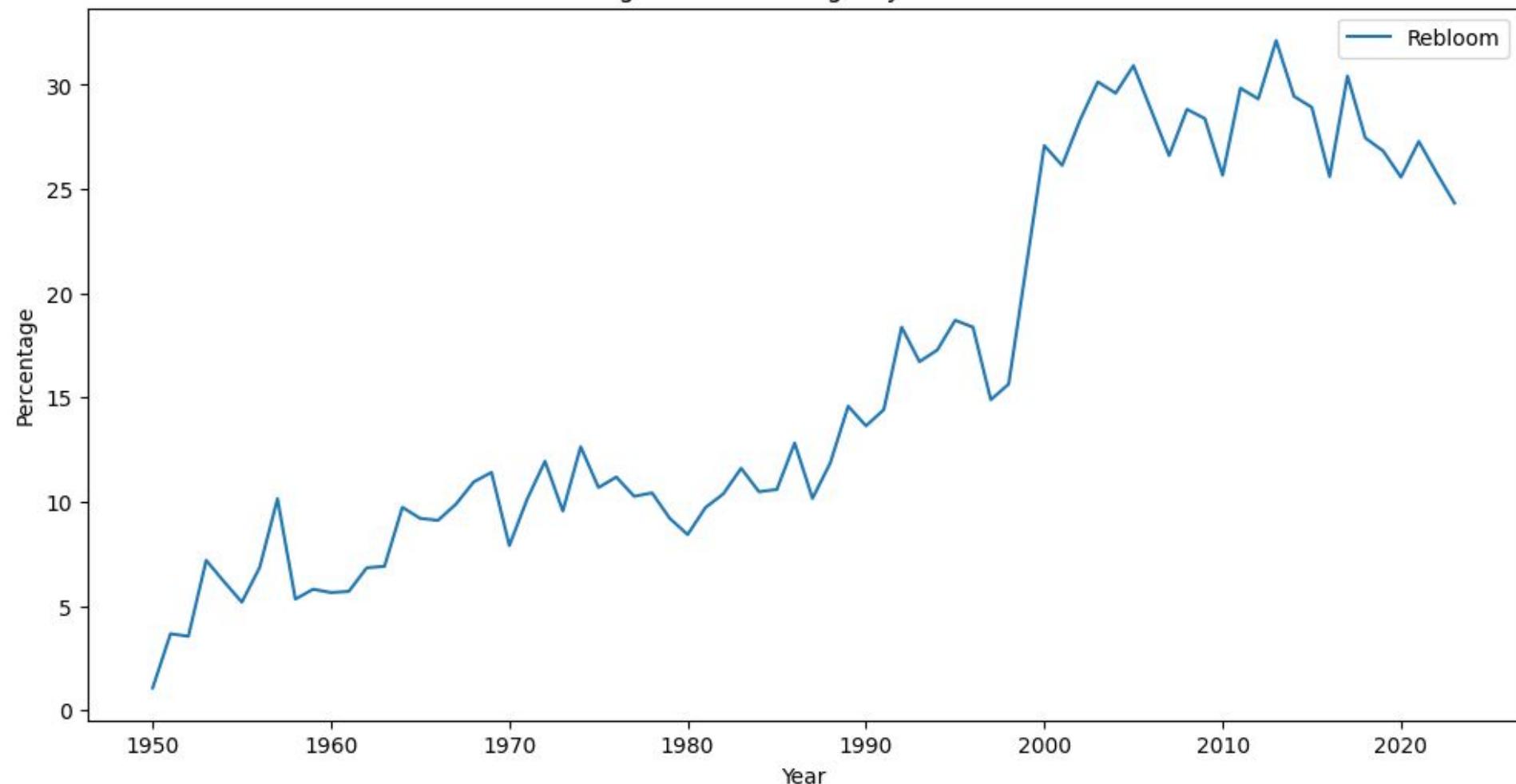
Percentage of Bloom Season Types Over Time



REBLOOM!!!!!!



Percentage of Reblooming Daylilies Over Time

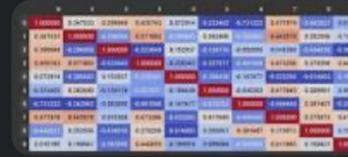


CORRELATION MATRIX

(Ewwww science words)

But what is it?

A correlation matrix is a statistical technique used to evaluate the relationship between two variables in a data set. The matrix is a table in which every cell contains a correlation coefficient, where 1 is considered a strong relationship between variables, 0 a neutral relationship and -1 a not strong relationship.



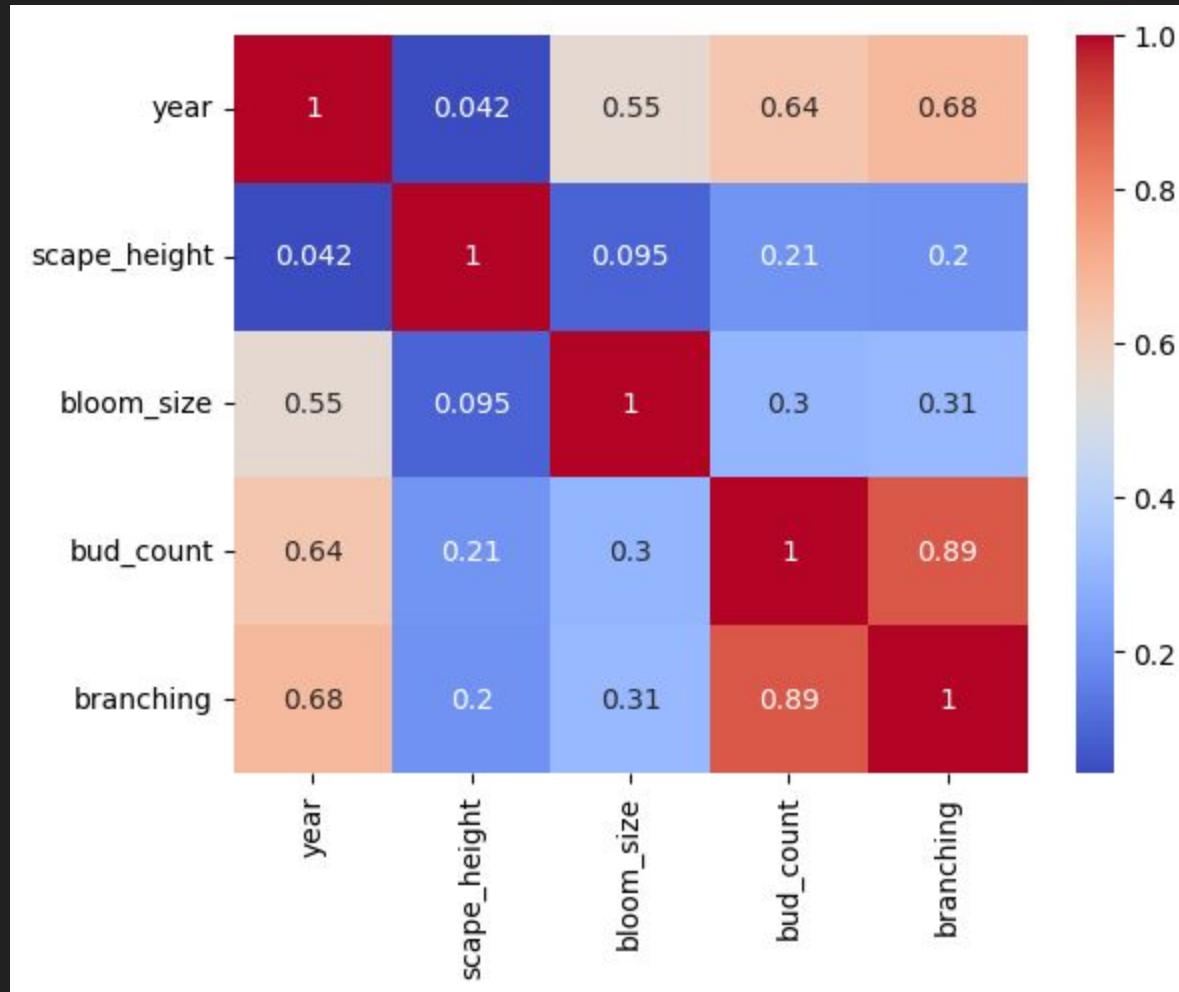
Mar 17, 2023



Built In

 <https://builtin.com> > data-science > correlation-matrix

Introduction to The Correlation Matrix | Built In





So what did we learn??

- Correlation between branching and budcount
- Mid season daylily is the most popular bloom season
- Dormant daylily are the most common type
- Tets are the most popular and common of recent intros (BOOOOOO)

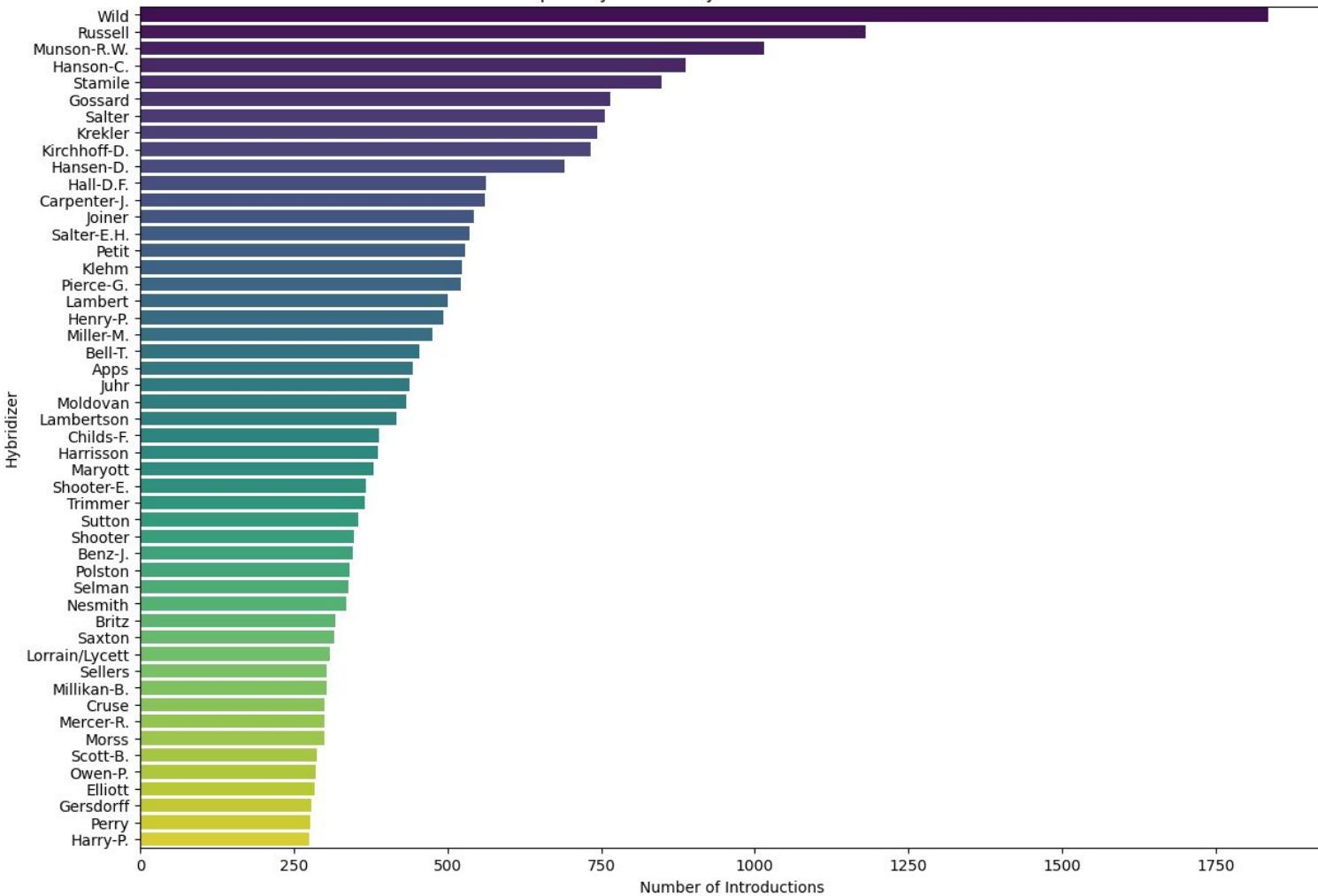
“Peppermint Puff” by Linda Donald

If we have all this data... who are the most prolific hybridizers...??? Most popular varieties to use for breeding???
Most unique genetics???

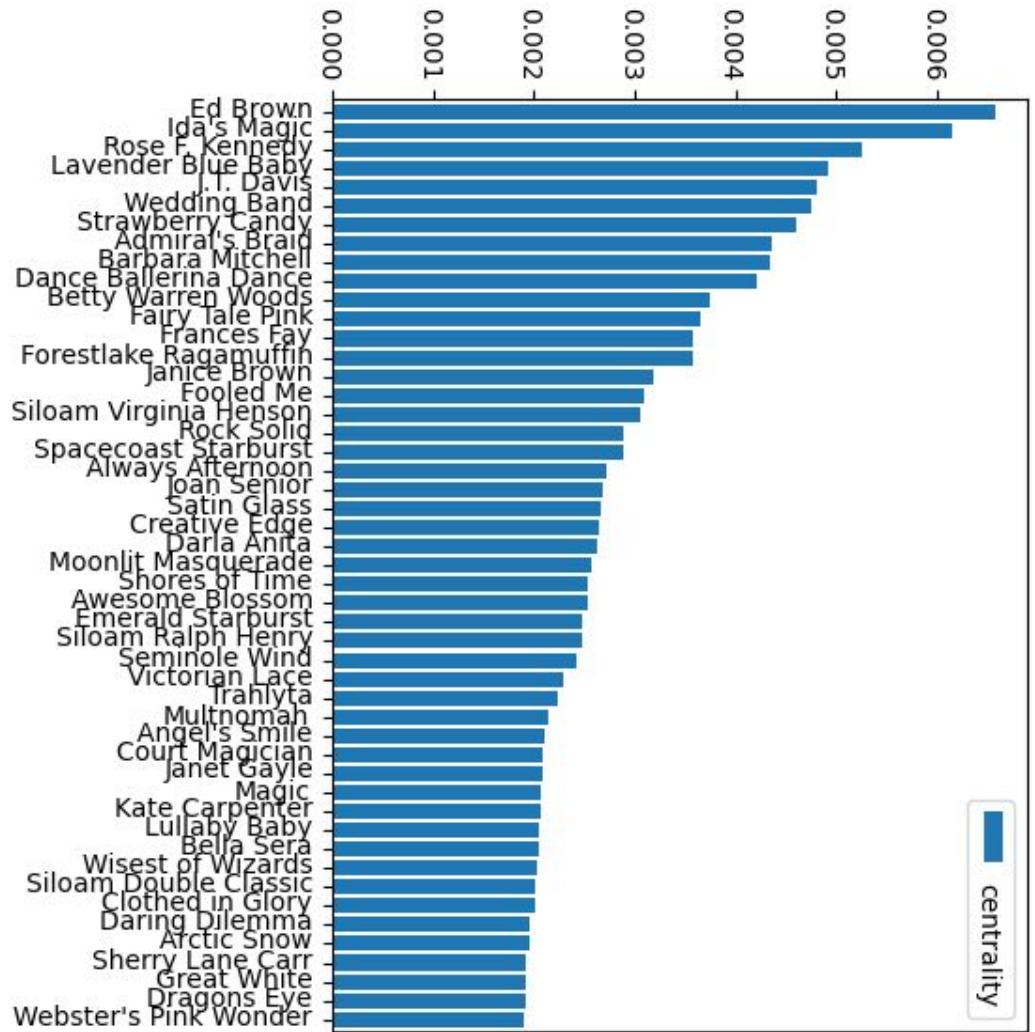
ARE THERE GROUPS OF DAYLILY???

Why stop here...

Top 50 Hybridizers by Number of Introductions



The Top 50



Most popular parents
for breeding as
measured by degree of
centrality

Last slide looks simple rightttttt????

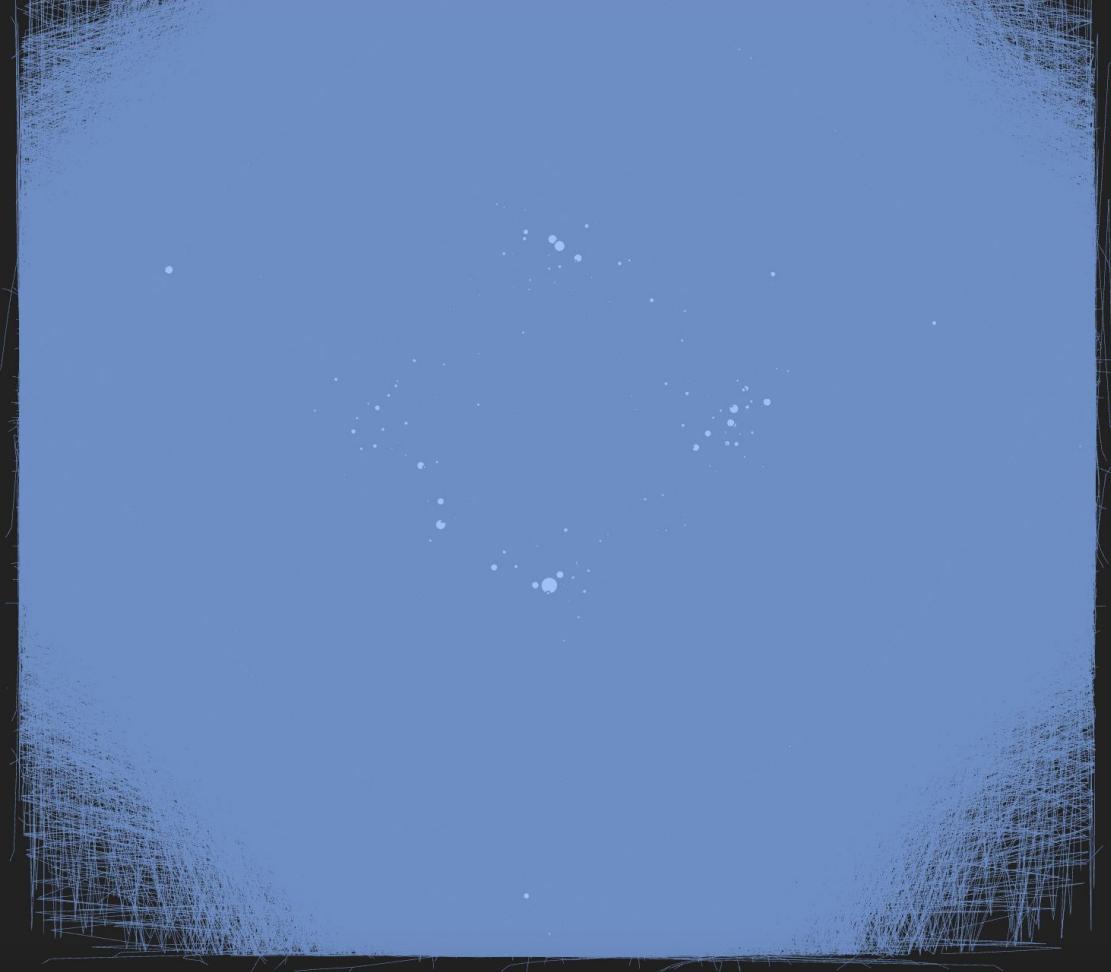


It took three months to figure out how to get the file where that algorithm could be applied and produce the centrality measures with over 40 versions of the code and even more swearing

Since this
source target
list is basically
a graph lets
see wut it
looks like

It took three days to load 😢

FOR THISSSSSS - >

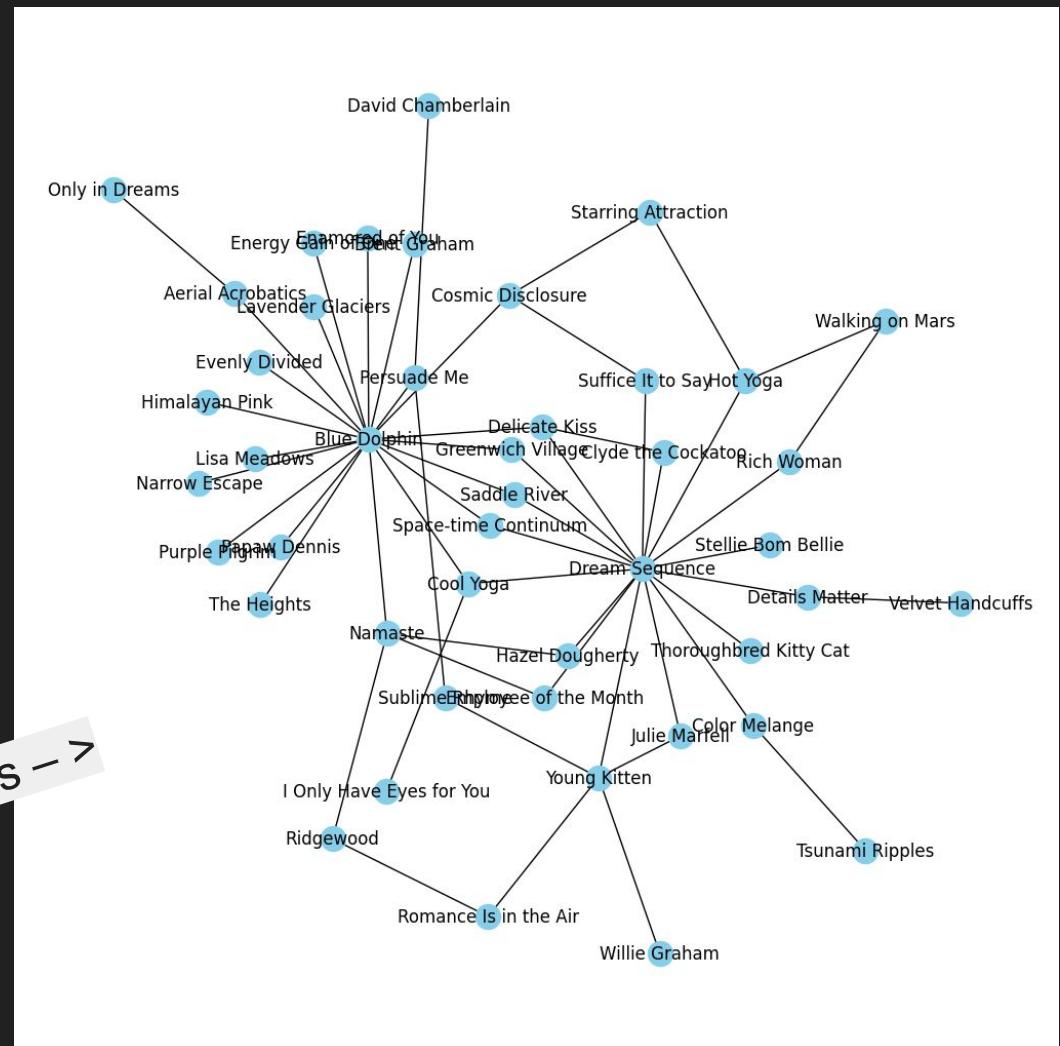


Subsets

Possible application

- Key stone varieties
- Ability to map unfamiliar programs
- Helps pick which varieties to buy

Rice's recent dip intros ->



My Goals for the future

- Sort out the rats nest
- Community detection?
- Map sub groups
- Look for useable varieties inside these groups to use in my own program by utilizing analytical methods
- Measure the impact of specific varieties to the genetic pool of daylilies as a whole



Q's?