

Predicting Pumpkin Weight



Connor, Ellann, Kim
& Emmy

Predict pumpkin weight



... because the **heaviest** pumpkin wins

AG & SQUASH COMBINED:

- 1st: \$6,000
- 2nd: \$3,000
- 3rd: \$2,000
- 4th: \$1,500
- 5th: \$1,100
- 6th: \$900
- 7th: \$800
- 8th: \$700
- 9th: \$600
- 10th: \$500
- 11th: \$300
- 12th: \$200
- 13th - 15th: \$100

Source: Bauman's Giant Pumpkin Weigh Off

How to grow a Giant Pumpkin:

Good Genetics

- Good seed
- Good pollen
- Good luck (runts happen!)

Good Grower (Skills)

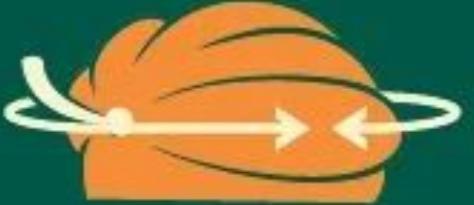
- Soil quality (fertilizer, soil acidity, & compost)
- Watering schedule
- Work your ass off (mulch, picking off the bugs)

Good Environment

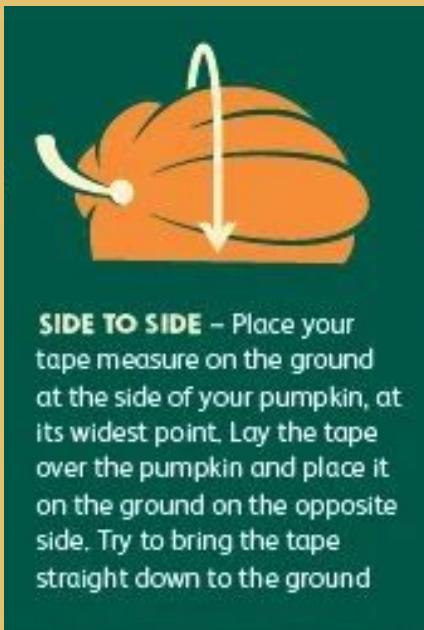
- Humidity
- Sunlight
- Temperature



Estimate the weight with OTT measurements



CIRCUMFERENCE – Measure all the way around your pumpkin at it's widest point, keeping your tape measure parallel with the ground

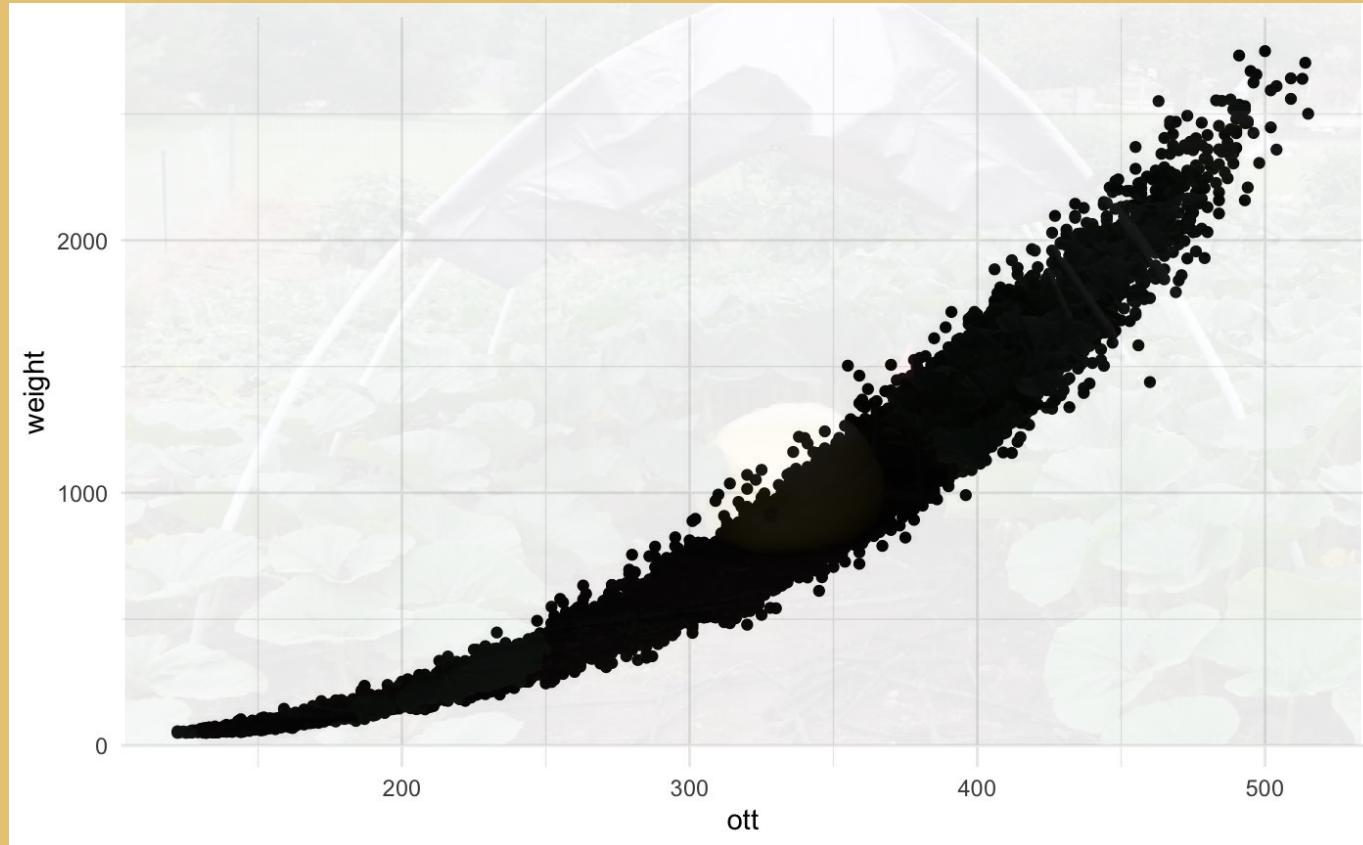


SIDE TO SIDE – Place your tape measure on the ground at the side of your pumpkin, at its widest point. Lay the tape over the pumpkin and place it on the ground on the opposite side. Try to bring the tape straight down to the ground



STEM TO BASE – Place your tape measure on the ground by the pumpkin stem. Lay it over the fruit, to the ground at the blossom end, or base and read off the measurement

Estimate the weight with OTT measurements



It's looks
pretty related

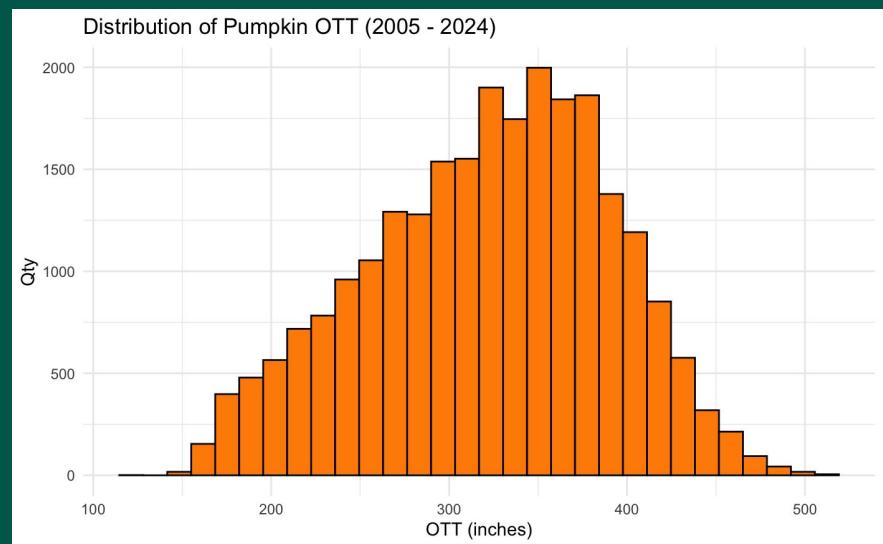
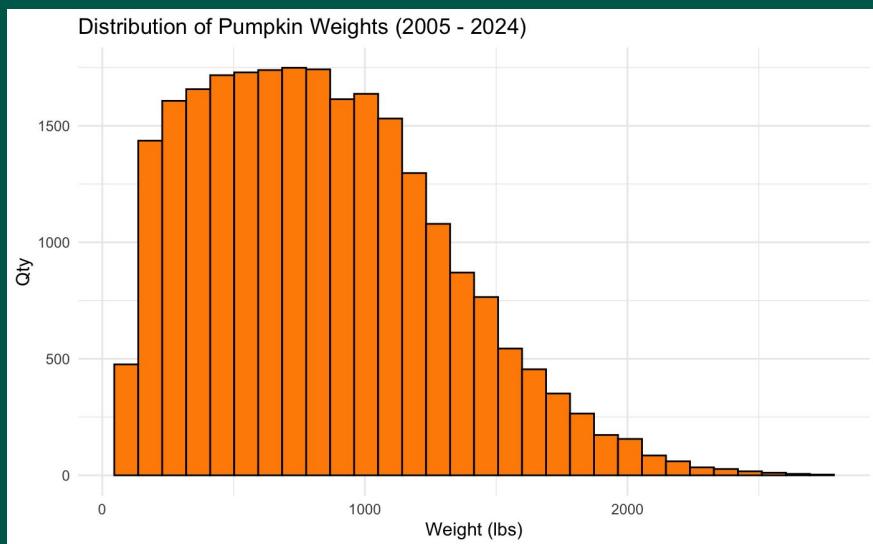
Pumpkin Patterns



... selective breeding means they're evolving!

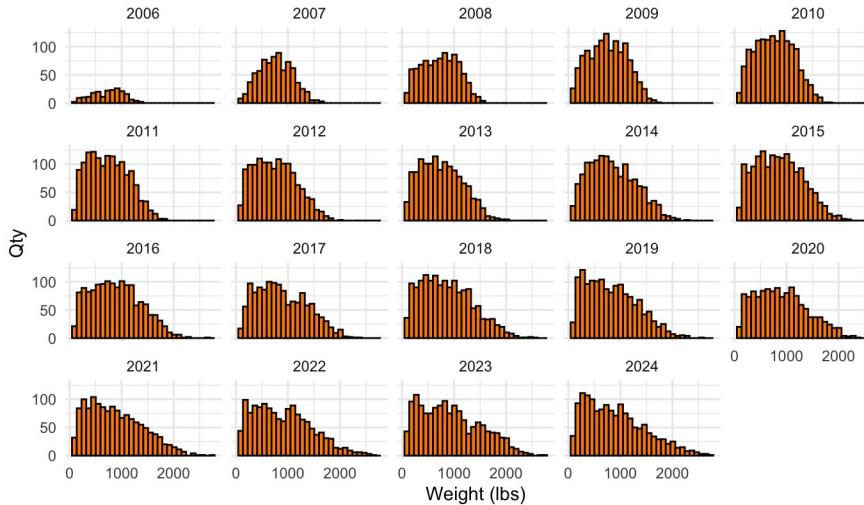


All the Pumpkins...





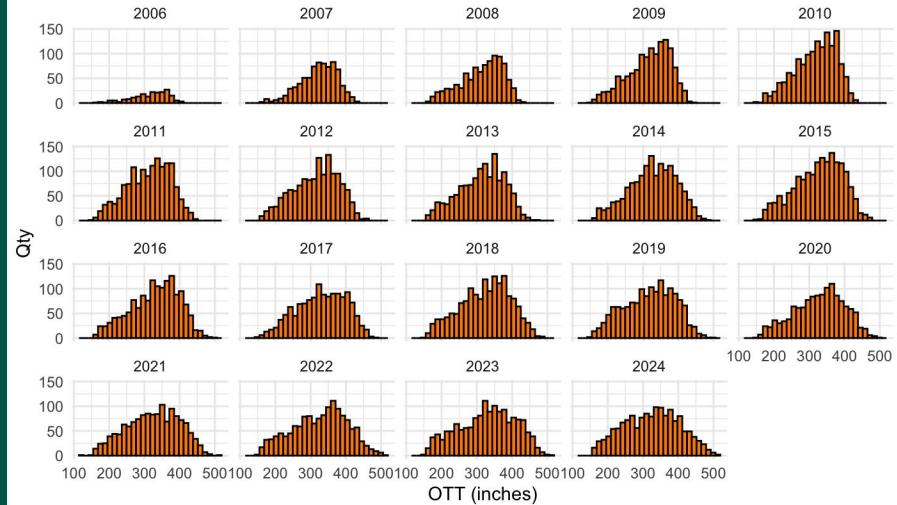
Distribution of Pumpkin Weights (2005 - 2024)



... are evolving!



Distribution of Pumpkin OTT (2005 - 2024)



Pumpkin Predictions



... can we perform better than the current formula?

How do growers estimate their pumpkin's weight?

- Use the lookup table based on this formula that uses OTT

2025 GPC OTT Chart for Atlantic Giant Pumpkins (inches to lbs)															
OTT	Ibs	OTT	Ibs	OTT	Ibs	OTT	Ibs	OTT	Ibs	OTT	Ibs	OTT	Ibs	OTT	Ibs
160	98	244	341	300	634	330	840	360	1080	390	1351	420	1651	450	1975
165	107	246	349	301	640	331	848	361	1088	391	1361	421	1661	451	1987
170	117	248	358	302	647	332	855	362	1097	392	1370	422	1672	452	1998
175	127	250	367	303	653	333	863	363	1106	393	1380	423	1682	453	2009
180	138	252	376	304	659	334	870	364	1114	394	1390	424	1693	454	2020
185	149	254	385	305	666	335	878	365	1123	395	1399	425	1703	455	2032
190	162	256	394	306	672	336	885	366	1132	396	1409	426	1714	456	2043
195	174	258	403	307	679	337	893	367	1140	397	1419	427	1724	457	2055
200	188	260	413	308	686	338	901	368	1149	398	1428	428	1735	458	2066
202	194	262	422	309	692	339	909	369	1158	399	1438	429	1746	459	2077
204	199	264	432	310	699	340	916	370	1167	400	1448	430	1756	460	2085
206	205	266	442	311	706	341	924	371	1176	401	1458	431	1767	461	2100
208	211	268	452	312	712	342	932	372	1185	402	1468	432	1778	462	2112
210	217	270	462	313	719	343	940	373	1194	403	1478	433	1789	463	2123
212	224	272	473	314	726	344	948	374	1203	404	1488	434	1799	464	2135
214	230	274	483	315	733	345	956	375	1212	405	1498	435	1810	465	2146
216	236	276	494	316	740	346	964	376	1221	406	1508	436	1821	466	2158
218	243	278	505	317	747	347	972	377	1230	407	1518	437	1832	467	2169
220	250	280	516	318	753	348	980	378	1239	408	1528	438	1843	468	2181
222	257	282	527	319	760	349	988	379	1248	409	1539	439	1854	469	2193
224	264	284	538	320	767	350	996	380	1257	410	1549	440	1864	470	2204
226	271	286	550	321	774	351	1005	381	1266	411	1558	441	1874	471	2214
228	278	288	561	322	782	352	1013	382	1275	412	1568	442	1883	472	2224
230	285	290	573	323	789	353	1021	383	1283	413	1578	443	1892	473	2234
232	293	292	585	324	796	354	1029	384	1292	414	1589	444	1901	474	2243
234	300	294	597	325	803	355	1037	385	1300	415	1598	445	1910	475	2252
238	316	297	615	326	810	356	1045	386	1309	416	1609	446	1919	476	2261
240	324	298	621	327	817	357	1053	387	1317	417	1618	447	1928	477	2270
242	332	299	628	328	824	358	1061	388	1326	418	1628	448	1937	478	2279

Weight_lbs = ((12.81/(1+6.87*

$$\left[\frac{12.81}{1 + 6.87 \cdot 2 \left(-\frac{OTT}{97} \right)} \right]^3 + \left(\frac{OTT}{45.9} \right)^{3.014} - 10$$

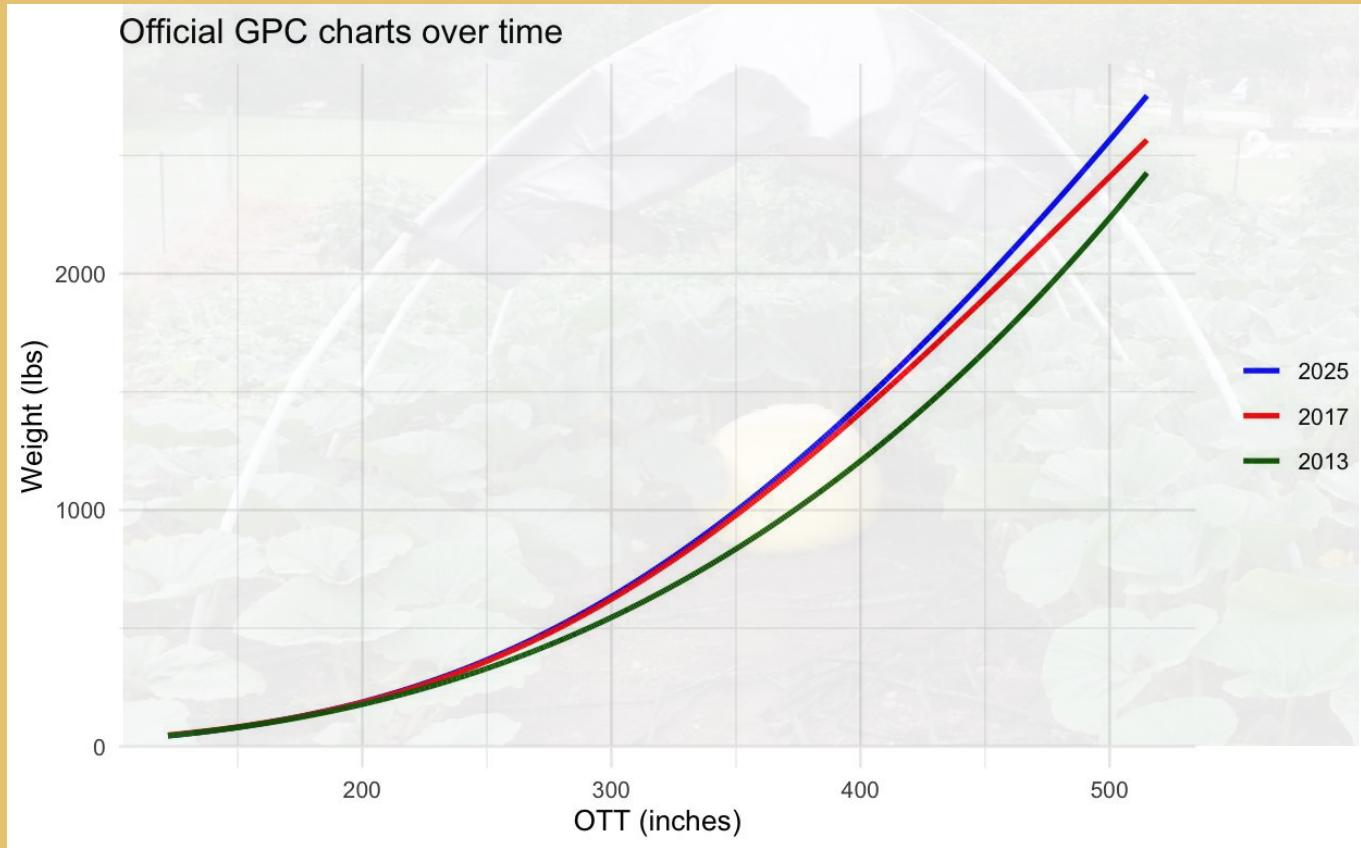
Why is that difficult?

... they're evolving!

- Competitive growers **selectively breed** pumpkins for desirable traits (maximal weight)
- Desirable traits include **increased fruit density**, so the formula changes periodically



Pumpkin estimations aren't evergreen



It looks like
it's evolving

Pumpkins



... transforming data into features

What data do we have?

Place	Weight (lbs)	Grower Name	City	State/Prov	Country	Seed (Mother)	Pollinator (Father)	OTT	Est. Weight	Pct. Chart
1	1,506.00	Cardwell, Harvey		Oregon	United States	1790.5 Wallace	sib	384.0	1,266.00	19.0
2	1,334.00	Sestak, Anne	Sublimity	Oregon	United States	2144 Rodebaugh	2069 Stelts	401.0	1,421.00	-6.0
3	1,230.50	Cohen, Ellann		Oregon	United States	1997.5 Sherwood	Self	404.0	1,449.00	-15.0
EXH	1,145.00	Cardwell, Harvey		Oregon	United States	1884 Tobeck	Self	361.0	1,068.00	7.0
EXH	1,120.50	Ford, Iris		Oregon	United States	1940 Urena	2469 Daletas	348.0	963.00	16.0
4	1,117.50	Ford, Iris		Oregon	United States	2294.5 Noel	2469 Daletas	364.0	1,093.00	2.0
5	1,088.50	Pugh, Edson	Madras	Oregon	United States	1381 Pugh 2023	1273 Pugh 1381	360.0	1,060.00	3.0
6	1,074.50	Spielman, Kendall	Brooks	Oregon	United States	2227 Stelts	1853 Ford	351.0	986.00	9.0
7	1,031.50	Cowger, Tobias	RIDGEFIELD	Washington	United States	2081 Werner	2088 Caspers	354.0	1,011.00	2.0
EXH	1,020.50	Pugh, Edson	Madras	Oregon	United States	1273 Pugh 2023	1381 Pugh 2023	356.0	1,027.00	-1.0
EXH	967.50	Sestak, Annie	Sublimity	Oregon	United States	1485 Sestak	2069 Stelts	353.0	1,002.00	-3.0
8	922.50	Guy, Gregory	Dallas	Oregon	United States	1953 Daletas	Open	339.5	897.00	3.0
9	885.50	Kristensen, Gary	Happy Valley	Oregon	United States	500 Est. Kristensen 2023	1296 Engle	322.0	769.00	15.0
10	858.00	Wollman, Liesl	Beaverton	Oregon	United States	1273 Ed Pugh	2501 Bernstrom	345.0	939.00	-9.0
11	806.00	Edwards, Ron	Dallas	Oregon	United States	1274 Pugh	Self	314.0	714.00	13.0
EXH	757.50	Spielman, Kendall	Brooks	Oregon	United States	1747.5 Handy		315.0	720.00	5.0
12	745.00	Possert, Jr, Michael	Warren	Oregon	United States	1796.5 Gadberry	Self	327.0	804.00	-7.0
13	658.00	Williams, Brian	Lake Oswego	Oregon	United States	1461 Rodebaugh	2465.5 Sperry	301.0	629.00	5.0
14	575.50	Williams, Luke	Portland	Oregon	United States	1997.5 Sherwood		302.0	636.00	-10.0
15	572.00	Farrester, Josh		Oregon	United States	1273 Pugh	Open	283.0	523.00	9.0
16	520.00	Bahns, Brad	Sherwood	Oregon	United States	2497 Kennedy Root	Open	272.0	464.00	12.0
EXH	515.00	Cardwell, Harvey		Oregon	United States	1798 Wallace	1790 Cardwell	279.0	501.00	3.0
17	506.00	Seklreta, Matt	Medford	Oregon	United States	1064 Kristensen	Self	286.0	540.00	-6.0
18	256.00	Sherwood, Jim	Mulino	Oregon	United States	2069 Stelts	Self	203.0	194.00	32.0
19	161.00	Clayton, Brandon	Glenoma	Washington	United States	527 Clayton	Sibling	188.0	155.00	4.0
EXH	156.00	Possert, Michael	Warren	Oregon	United States	466 Possert	1796.5 Gadberry	186.0	151.00	3.0
DMG	1,143.50	Frankfort, John		Oregon	United States			372.0	1,161.00	-2.0
DMG	806.50	Barker, Ron	BONNEY LAKE	Washington	United States	2501 Bernstrom	1997.5 Sherwood	334.0	855.00	-6.0

28 Entries. (7 were exhibition only.)

More than a
field-full with
30k pumpkins



~25k after cleanup

Source: BigPumpkins.com
Shown: 2024 Tualatin Results

Goal: Predict the Weight

Place	Weight (lbs)	Grower Name	City	State/Prov	Country	Seed (Mother)	Pollinator (Father)	OTT	Est. Weight	Pct. Chart
1	1,506.00	Cardwell, Harvey		Oregon	United States	1790.5 Wallace	sib	384.0	1,266.00	19.0
2	1,334.00	Sestak, Anne	Sublimity	Oregon	United States	2144 Rodebaugh	2069 Stelts	401.0	1,421.00	-6.0
3	1,230.50	Cohen, Ellann		Oregon	United States	1997.5 Sherwood	Self	404.0	1,449.00	-15.0
EXH	1,145.00	Cardwell, Harvey		Oregon	United States	1884 Tobeck	Self	361.0	1,068.00	7.0
EXH	1,120.50	Ford, Iris		Oregon	United States	1940 Urena	2469 Daletas	348.0	963.00	16.0
4	1,117.50	Ford, Iris		Oregon	United States	2294.5 Noel	2469 Daletas	364.0	1,093.00	2.0
5	1,088.50	Pugh, Edson	Madras	Oregon	United States	1381 Pugh 2023	1273 Pugh 1381	360.0	1,060.00	3.0
6	1,074.50	Spielman, Kendall	Brooks	Oregon	United States	2227 Stelts	1853 Ford	351.0	986.00	9.0
7	1,031.50	Cowger, Tobias	RIDGEFIELD	Washington	United States	2081 Werner	2088 Caspers	354.0	1,011.00	2.0
EXH	1,020.50	Pugh, Edson	Madras	Oregon	United States	1273 Pugh 2023	1381 Pugh 2023	356.0	1,027.00	-1.0
EXH	967.50	Sestak, Annie	Sublimity	Oregon	United States	1485 Sestak	2069 Stelts	353.0	1,002.00	-3.0
8	922.50	Guy, Gregory	Dallas	Oregon	United States	1953 Daletas	Open	339.5	897.00	3.0
9	885.50	Kristensen, Gary	Happy Valley	Oregon	United States	500 Est. Kristensen 2023	1296 Engle	322.0	769.00	15.0
10	858.00	Wollman, Liesl	Beaverton	Oregon	United States	1273 Ed Pugh	2501 Bernstrom	345.0	939.00	-9.0
11	806.00	Edwards, Ron	Dallas	Oregon	United States	1274 Pugh	Self	314.0	714.00	13.0
EXH	757.50	Spielman, Kendall	Brooks	Oregon	United States	1747.5 Handy		315.0	720.00	5.0
12	745.00	Possert, Jr, Michael	Warren	Oregon	United States	1796.5 Gadberry	Self	327.0	804.00	-7.0
13	658.00	Williams, Brian	Lake Oswego	Oregon	United States	1461 Rodebaugh	2465.5 Sperry	301.0	629.00	5.0
14	575.50	Williams, Luke	Portland	Oregon	United States	1997.5 Sherwood		302.0	636.00	-10.0
15	572.00	Farrester, Josh		Oregon	United States	1273 Pugh	Open	283.0	523.00	9.0
16	520.00	Bahns, Brad	Sherwood	Oregon	United States	2497 Kennedy Root	Open	272.0	464.00	12.0
EXH	515.00	Cardwell, Harvey		Oregon	United States	1798 Wallace	1790 Cardwell	279.0	501.00	3.0
17	506.00	Seklreta, Matt	Medford	Oregon	United States	1064 Kristensen	Self	286.0	540.00	-6.0
18	256.00	Sherwood, Jim	Mulino	Oregon	United States	2069 Stelts	Self	203.0	194.00	32.0
19	161.00	Clayton, Brandon	Glenoma	Washington	United States	527 Clayton	Sibling	188.0	155.00	4.0
EXH	156.00	Possert, Michael	Warren	Oregon	United States	466 Possert	1796.5 Gadberry	186.0	151.00	3.0
DMG	1,143.50	Frankfort, John		Oregon	United States			372.0	1,161.00	-2.0
DMG	806.50	Barker, Ron	BONNEY LAKE	Washington	United States	2501 Bernstrom	1997.5 Sherwood	334.0	855.00	-6.0

28 Entries. (7 were exhibition only.)



Current Method: OTT Only



Place	Weight (lbs)	Grower Name	City	State/Prov	Country	Seed (Mother)	Pollinator (Father)	OTT	Est. Weight	Pct. Chart
1	1,506.00	Cardwell, Harvey		Oregon	United States	1790.5 Wallace	sib	384.0	1,266.00	19.0
2	1,334.00	Sestak, Anne	Sublimity	Oregon	United States	2144 Rodebaugh	2069 Stelts	401.0	1,421.00	-6.0
3	1,230.50	Cohen, Ellann		Oregon	United States	1997.5 Sherwood	Self	404.0	1,449.00	-1.0
EXH	1,145.00	Cardwell, Harvey		Oregon	United States	1884 Tobeck	Self	361.0	1,068.00	
EXH	1,120.50	Ford, Iris		Oregon	United States	1940 Urena	2469 Daletas	348.0	963.00	
4	1,117.50	Ford, Iris		Oregon	United States	2294.5 Noel	2469 Daletas	364.0	1,093.00	
5	1,088.50	Pugh, Edson	Madras	Oregon	United States	1381 Pugh 2023	1273 Pugh 1381	360.0	1,060.00	
6	1,074.50	Spielman, Kendall	Brooks	Oregon	United States	2227 Stelts	1853 Ford	351.0	986.00	9.0
7	1,031.50	Cowger, Tobias	RIDGEFIELD	Washington	United States	2081 Werner	2088 Caspers	354.0	1,011.00	2.0
EXH	1,020.50	Pugh, Edson	Madras	Oregon	United States	1273 Pugh 2023	1381 Pugh 2023	356.0	1,027.00	-1.0
EXH	967.50	Sestak, Annie	Sublimity	Oregon	United States	1485 Sestak	2069 Stelts	353.0	1,002.00	-3.0
8	922.50	Guy, Gregory	Dallas	Oregon	United States	1953 Daletas	Open	339.5	897.00	3.0
9	885.50	Kristensen, Gary	Happy Valley	Oregon	United States	500 Est. Kristensen 2023	1296 Engle	322.0	769.00	15.0
10	858.00	Wollman, Liesl	Beaverton	Oregon	United States	1273 Ed Pugh	2501 Bernstrom	345.0	939.00	-9.0
11	806.00	Edwards, Ron	Dallas	Oregon	United States	1274 Pugh	Self	314.0	714.00	13.0
EXH	757.50	Spielman, Kendall	Brooks	Oregon	United States	1747.5 Handy		315.0	720.00	5.0
12	745.00	Possert, Jr, Michael	Warren	Oregon	United States	1796.5 Gadberry	Self	327.0	804.00	-7.0
13	658.00	Williams, Brian	Lake Oswego	Oregon	United States	1461 Rodebaugh	2465.5 Sperry	301.0	629.00	5.0
14	575.50	Williams, Luke	Portland	Oregon	United States	1997.5 Sherwood		302.0	636.00	-10.0
15	572.00	Farrester, Josh		Oregon	United States	1273 Pugh	Open	283.0	523.00	9.0
16	520.00	Bahns, Brad	Sherwood	Oregon	United States	2497 Kennedy Root	Open	272.0	464.00	12.0
EXH	515.00	Cardwell, Harvey		Oregon	United States	1798 Wallace	1790 Cardwell	279.0	501.00	3.0
17	506.00	Seklreta, Matt	Medford	Oregon	United States	1064 Kristensen	Self	286.0	540.00	-6.0
18	256.00	Sherwood, Jim	Mulino	Oregon	United States	2069 Stelts	Self	203.0	194.00	32.0
19	161.00	Clayton, Brandon	Glenoma	Washington	United States	527 Clayton	Sibling	188.0	155.00	4.0
EXH	156.00	Possert, Michael	Warren	Oregon	United States	466 Possert	1796.5 Gadberry	186.0	151.00	3.0
DMG	1,143.50	Frankfort, John		Oregon	United States			372.0	1,161.00	-2.0
DMG	806.50	Barker, Ron	BONNEY LAKE	Washington	United States	2501 Bernstrom	1997.5 Sherwood	334.0	855.00	-6.0

28 Entries. (7 were exhibition only.)

$$\left(\frac{12.81}{1 + 6.87 * 2} \left(\frac{-OTT}{97} \right) \right)^3 + \left(\frac{OTT}{45.9} \right)^{3.014} - 10$$

Current Method: Accurate?



Place	Weight (lbs)	Grower Name	City	State/Prov	Country	Seed (Mother)	Pollinator (Father)	OTT	Est. Weight	Pct. Chart
1	1,506.00	Cardwell, Harvey		Oregon	United States	1790.5 Wallace	sib	384.0	1,266.00	19.0
2	1,334.00	Sestak, Anne	Sublimity	Oregon	United States	2144 Rodebaugh	2069 Stelts	401.0	1,421.00	-6.0
3	1,230.50	Cohen, Ellann		Oregon	United States	1997.5 Sherwood	Self	404.0	1,449.00	-15.0
EXH	1,145.00	Cardwell, Harvey		Oregon	United States	1884 Tobeck	Self	361.0	1,068.00	7.0
EXH	1,120.50	Ford, Iris		Oregon	United States	1940 Urena	2469 Daletas	348.0	963.00	16.0
4	1,117.50	Ford, Iris		Oregon	United States	2294.5 Noel	2469 Daletas	364.0	1,093.00	2.0
5	1,088.50	Pugh, Edson	Madras	Oregon	United States	1381 Pugh 2023	1273 Pugh 1381	360.0	1,060.00	3.0
6	1,074.50	Spielman, Kendall	Brooks	Oregon	United States	2227 Stelts	1853 Ford	351.0	986.00	9.0
7	1,031.50	Cowger, Tobias	RIDGEFIELD	Washington	United States	2081 Werner	2088 Caspers	354.0	1,011.00	2.0
EXH	1,020.50	Pugh, Edson	Madras	Oregon	United States	1273 Pugh 2023	1381 Pugh 2023	356.0	1,027.00	-1.0
EXH	967.50	Sestak, Annie	Sublimity	Oregon	United States	1485 Sestak	2069 Stelts	353.0	1,002.00	-3.0
8	922.50	Guy, Gregory	Dallas	Oregon	United States	1953 Daletas	Open	339.5	897.00	3.0
9	885.50	Kristensen, Gary	Happy Valley	Oregon	United States	500 Est. Kristensen 2023	1296 Engle	322.0	769.00	15.0
10	858.00	Wollman, Liesl	Beaverton	Oregon	United States	1273 Ed Pugh	2501 Bernstrom	345.0	939.00	-9.0
11	806.00	Edwards, Ron	Dallas	Oregon	United States	1274 Pugh	Self	314.0	714.00	13.0
EXH	757.50	Spielman, Kendall	Brooks	Oregon	United States	1747.5 Handy		315.0	720.00	5.0
12	745.00	Possert, Jr, Michael	Warren	Oregon	United States	1796.5 Gadberry	Self	327.0	804.00	-7.0
13	658.00	Williams, Brian	Lake Oswego	Oregon	United States	1461 Rodebaugh	2465.5 Sperry	301.0	629.00	5.0
14	575.50	Williams, Luke	Portland	Oregon	United States	1997.5 Sherwood		302.0	636.00	-10.0
15	572.00	Farrester, Josh		Oregon	United States	1273 Pugh	Open	283.0	523.00	9.0
16	520.00	Bahns, Brad	Sherwood	Oregon	United States	2497 Kennedy Root	Open	272.0	464.00	12.0
EXH	515.00	Cardwell, Harvey		Oregon	United States	1798 Wallace	1790 Cardwell	279.0	501.00	3.0
17	506.00	Seklreta, Matt	Medford	Oregon	United States	1064 Kristensen	Self	286.0	540.00	-6.0
18	256.00	Sherwood, Jim	Mulino	Oregon	United States	2069 Stelts	Self	203.0	194.00	32.0
19	161.00	Clayton, Brandon	Glenoma	Washington	United States	527 Clayton	Sibling	188.0	155.00	4.0
EXH	156.00	Possert, Michael	Warren	Oregon	United States	466 Possert	1796.5 Gadberry	186.0	151.00	3.0
DMG	1,143.50	Frankfort, John		Oregon	United States			372.0	1,161.00	-2.0
DMG	806.50	Barker, Ron	BONNEY LAKE	Washington	United States	2501 Bernstrom	1997.5 Sherwood	334.0	855.00	-6.0

28 Entries. (7 were exhibition only.)

$$\left(\frac{12.81}{1 + 6.87*2} \left(\frac{-OTT}{97} \right) \right)^3 + \left(\frac{OTT}{45.9} \right)^{3.014} - 10$$

Source: BigPumpkins.com
Shown: 2024 Tualatin Results

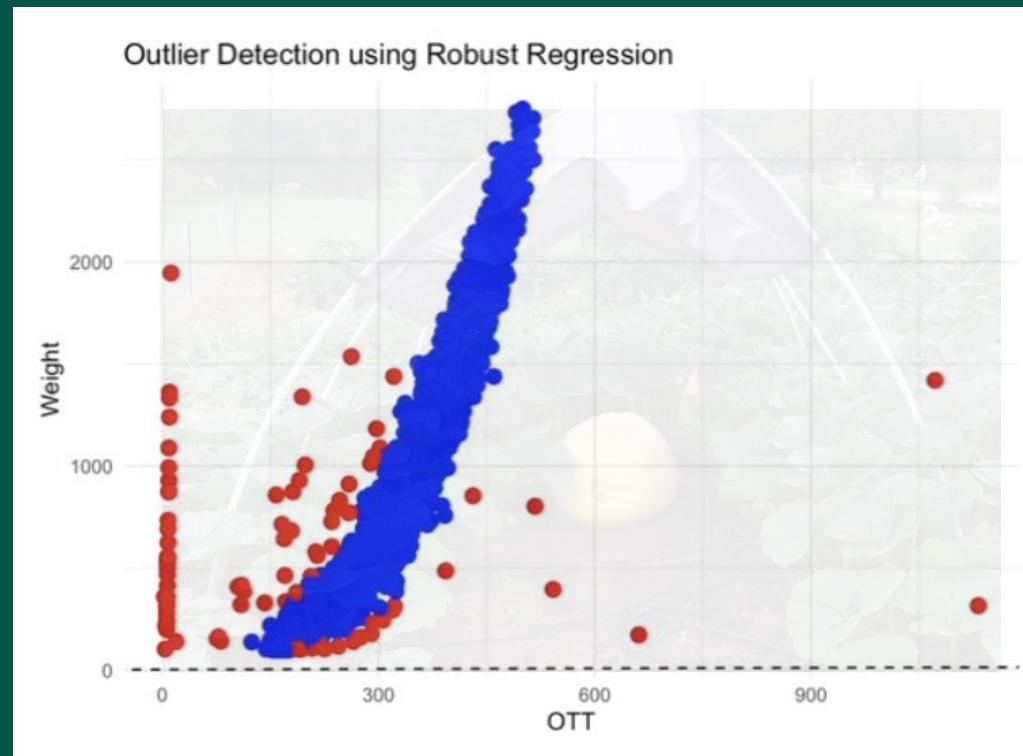
Data Cleaning

Robust Linear Regression with Log-Transformation and Residual-Based Outlier Filtering

Robust Linear Model (RLM) with `r1m()`
Downweights outliers during fitting, resists being pulled by typos.

Log-transform the weight ($\log(\text{weight})$)
Weight increases exponentially with OTT at higher sizes, *this linearizes* the relationship.

Outlier detection based on residuals
Flags pumpkins that deviate from the model as outliers and filters them out.



Pumpkin Parents (New Features)

Place	Weight (lbs)	Grower Name	City	State/Prov	Country	Seed (Mother)	Pollinator (Father)	OTT	Est. Weight	Pct. Chart
1	1,506.00	Cardwell, Harvey		Oregon	United States	1790.5 Wallace	sib	384.0	1,266.00	19.0
2	1,334.00	Sestak, Anne	Sublimity	Oregon	United States	2144 Rodebaugh	2069 Stelts	401.0	1,421.00	-6.0
3	1,230.50	Cohen, Ellann		Oregon	United States	1997.5 Sherwood	Self	404.0	1,449.00	-15.0
EXH	1,145.00	Cardwell, Harvey		Oregon	United States	1884 Tobeck	Self	361.0	1,068.00	7.0
EXH	1,120.50	Ford, Iris		Oregon	United States	1940 Urena	2469 Daletas	348.0	963.00	16.0
4	1,117.50	Ford, Iris		Oregon	United States	2294.5 Noel	2469 Daletas	364.0	1,093.00	2.0
5	1,088.50	Pugh, Edson	Madras	Oregon	United States	1381 Pugh 2023	1273 Pugh 1381	360.0	1,060.00	3.0
6	1,074.50	Spielman, Kendall	Brooks	Oregon	United States	2227 Stelts	1853 Ford	351.0	986.00	9.0
7	1,031.50	Cowger, Tobias	RIDGEFIELD	Washington	United States	2081 Werner	2088 Caspers	354.0	1,011.00	2.0
EXH	1,020.50	Pugh, Edson	Madras	Oregon	United States	1273 Pugh 2023	1381 Pugh 2023	356.0	1,027.00	-1.0
EXH	967.50	Sestak, Annie	Sublimity	Oregon	United States	1485 Sestak	2069 Stelts	353.0	1,002.00	-3.0
8	922.50	Guy, Gregory	Dallas	Oregon	United States	1953 Daletas	Open	339.5	897.00	3.0
9	885.50	Kristensen, Gary	Happy Valley	Oregon	United States	500 Est. Kristensen 2023	1296 Engle	322.0	769.00	15.0
10	858.00	Wollman, Liesl	Beaverton	Oregon	United States	1273 Ed Pugh	2501 Bernstrom	345.0	939.00	-9.0
11	806.00	Edwards, Ron	Dallas	Oregon	United States	1274 Pugh	Self	314.0	714.00	13.0
EXH	757.50	Spielman, Kendall	Brooks	Oregon	United States	1747.5 Handy		315.0	720.00	5.0
12	745.00	Possert, Jr, Michael	Warren	Oregon	United States	1796.5 Gadberry	Self	327.0	804.00	-7.0
13	658.00	Williams, Brian	Lake Oswego	Oregon	United States	1461 Rodebaugh	2465.5 Sperry	301.0	629.00	5.0
14	575.50	Williams, Luke	Portland	Oregon	United States	1997.5 Sherwood		302.0	636.00	-10.0
15	572.00	Farrester, Josh		Oregon	United States	1273 Pugh	Open	283.0	523.00	9.0
16	520.00	Bahns, Brad	Sherwood	Oregon	United States	2497 Kennedy Root	Open	272.0	464.00	12.0
EXH	515.00	Cardwell, Harvey		Oregon	United States	1798 Wallace	1790 Cardwell	279.0	501.00	3.0
17	506.00	Seklreta, Matt	Medford	Oregon	United States	1064 Kristensen	Self	286.0	540.00	-6.0
18	256.00	Sherwood, Jim	Mulino	Oregon	United States	2069 Stelts	Self	203.0	194.00	32.0
19	161.00	Clayton, Brandon	Glenoma	Washington	United States	527 Clayton	Sibling	188.0	155.00	4.0
EXH	156.00	Possert, Michael	Warren	Oregon	United States	466 Possert	1796.5 Gadberry	186.0	151.00	3.0
DMG	1,143.50	Frankfort, John		Oregon	United States			372.0	1,161.00	-2.0
DMG	806.50	Barker, Ron	BONNEY LAKE	Washington	United States	2501 Bernstrom	1997.5 Sherwood	334.0	855.00	-6.0

28 Entries. (7 were exhibition only.)

Weight
+ Grower



Parsing



Pumpkin Parents (Data Cleaning)

Place	Weight (lbs)	Grower Name	City	State/Prov	Country	Seed (Mother)	Pollinator (Father)	OTT	Est. Weight	Pct. Chart
1	1,506.00	Cardwell, Harvey		Oregon	United States	1790.5 Wallace	sib	384.0	1,266.00	19.0
2	1,334.00	Sestak, Anne	Sublimity	Oregon	United States	2144 Rodebaugh	2069 Stelts	401.0	1,421.00	-6.0
3	1,230.50	Cohen, Ellann		Oregon	United States	1997.5 Sherwood	Self	404.0	1,449.00	-15.0
EXH	1,145.00	Cardwell, Harvey		Oregon	United States	1884 Tobeck	Self	361.0	1,068.00	7.0
EXH	1,120.50	Ford, Iris		Oregon	United States	1940 Urena	2469 Daletas	348.0	963.00	16.0
4	1,117.50	Ford, Iris		Oregon	United States	2294.5 Noel	2469 Daletas	364.0	1,093.00	2.0
5	1,088.50	Pugh, Edson	Madras	Oregon	United States	1381 Pugh 2023	1273 Pugh 1381	360.0	1,060.00	3.0
6	1,074.50	Spielman, Kendall	Brooks	Oregon	United States	2227 Stelts	1853 Ford	351.0	986.00	9.0
7	1,031.50	Cowger, Tobias	RIDGEFIELD	Washington	United States	2081 Werner	2088 Caspers	354.0	1,011.00	2.0
EXH	1,020.50	Pugh, Edson	Madras	Oregon	United States	1273 Pugh 2023	1381 Pugh 2023	356.0	1,027.00	-1.0
EXH	967.50	Sestak, Annie	Sublimity	Oregon	United States	1485 Sestak	2069 Stelts	353.0	1,002.00	-3.0
8	922.50	Guy, Gregory	Dallas	Oregon	United States	1953 Daletas	Open	339.5	897.00	3.0
9	885.50	Kristensen, Gary	Happy Valley	Oregon	United States	500 Est. Kristensen 2023	1296 Engle	322.0	769.00	15.0
10	858.00	Wollman, Liesl	Beaverton	Oregon	United States	1273 Ed Pugh	2501 Bernstrom	345.0	939.00	-9.0
11	806.00	Edwards, Ron	Dallas	Oregon	United States	1274 Pugh	Self	314.0	714.00	13.0
EXH	757.50	Spielman, Kendall	Brooks	Oregon	United States	1747.5 Handy		315.0	720.00	5.0
12	745.00	Possert, Jr, Michael	Warren	Oregon	United States	1796.5 Gadberry	Self	327.0	804.00	-7.0
13	658.00	Williams, Brian	Lake Oswego	Oregon	United States	1461 Rodebaugh	2465.5 Sperry	301.0	629.00	5.0
14	575.50	Williams, Luke			United States	1997.5 Sherwood		302.0	636.00	-10.0
15	572.00	Farrester, Josh			United States	1273 Pugh	Open	283.0	523.00	9.0
16	520.00	Bahns, Brad			United States	2497 Kennedy Root	Open	272.0	464.00	12.0
EXH	515.00	Cardwell, Harvey			United States	1798 Wallace	1790 Cardwell	279.0	501.00	3.0
17	506.00	Seklreta, Matt			United States	1064 Kristensen	Self	286.0	540.00	-6.0
18	256.00	Sherwood, Jim	Mulino	Oregon	United States	2069 Stelts	Self	203.0	194.00	32.0
19	161.00	Clayton, Brandon	Glenoma	Washington	United States	527 Clayton	Sibling	188.0	155.00	4.0
EXH	156.00	Possert, Michael	Warren	Oregon	United States	466 Possert	1796.5 Gadberry	186.0	151.00	3.0
DMG	1,143.50	Frankfort, John		Oregon	United States			372.0	1,161.00	-2.0
DMG	806.50	Barker, Ron	BONNEY LAKE	Washington	United States	2501 Bernstrom	1997.5 Sherwood	334.0	855.00	-6.0

28 Entries. (7 were exhibition only.)

Incomplete
& Unknown

Dirty Data



Pumpkin Parents (Data Cleaning)

Place	Weight (lbs)	Grower Name	City	Seed (Mother)	Weight (Mother)	Grower (Mother)	Weight Pct. Chart
1	1,506.00	Cardwell, Harvey	MADRAS	1790.5 Wallace	1790.5	Wallace	1,266.00 19.0
				2144 Rodebaugh	2144	Rodebaugh	1,421.00 -6.0
				1997.5 Sherwood	1997.5	Sherwood	1,449.00 -15.0
				1884 Tobeck	1884	Tobeck	1,068.00 7.0
5	1,088.50	Pugh, Edson	Madras				963.00 16.0
6	1,074.50	Spielman, Kendall	Brooks	Oregon	United States	2227 Stelts	1,093.00 2.0
7	1,031.50	Cowger, Tobias	RIDGEFIELD	Washington	United States	2081 Werner	1,060.00 3.0
EXH	1,020.50	Pugh, Edson	Madras	Oregon	United States	1273 Pugh 2023	1,027.00 -1.0
EXH	967.50	Sestak, Annie	Sublimity	Oregon	United States	1485 Sestak	1,002.00 -3.0
8	922.50	Guy, Gregory	Dallas	Oregon	United States	1953 Daletas	897.00 3.0
9	885.50	Kristensen, Gary	Happy Valley	Oregon	United States	500 Est. Kristensen 2023	769.00 15.0
10	858.00	Wollman, Liesl	Beaverton	Oregon	United States	1273 Ed Pugh	939.00 -9.0
11	806.00	Edwards, Ron	Dallas	Oregon	United States	1274 Pugh	714.00 13.0
EXH	757.50	Spielman, Kendall	Brooks	Oregon	United States	1747.5 Handy	720.00 5.0
12	745.00	Possert, Jr, Michael	Warren	Oregon	United States	1796.5 Gadberry	804.00 -7.0
13	658.00	Williams, Brian	Lake Oswego	Oregon	United States	1461 Rodebaugh	629.00 5.0
14	575.50	Williams, Luke	Portland	Oregon	United States	1997.5 Sherwood	636.00 -10.0
15	572.00	Farrester, Josh		Oregon	United States	1273 Pugh	523.00 9.0
16	520.00	Bahns, Brad	Sherwood	Oregon	United States	2497 Kennedy Root	464.00 12.0
EXH	515.00	Cardwell, Harvey		Oregon	United States	1798 Wallace	501.00 3.0
17	506.00	Seklreta, Matt	Medford	Oregon	United States	1064 Kristensen	540.00 -6.0
18	256.00	Sherwood, Jim	Mulino	Oregon	United States	2069 Stelts	194.00 32.0
19	161.00	Clayton, Brandon	Glenoma	Washington	United States	527 Clayton	155.00 4.0
EXH	156.00	Possert, Michael	Warren	Oregon	United States	466 Possert	151.00 3.0
DMG	1,143.50	Frankfort, John		Oregon	United States		1,161.00 -2.0
DMG	806.50	Barker, Ron	BONNEY LAKE	Washington	United States	2501 Bernstrom	855.00 -6.0

28 Entries. (7 were exhibition only.)

Dirty Data



Pumpkin Grower Location (New Feature)

Place	Weight (lbs)	Grower Name	City	State/Prov	Country	Seed (Mother)	Pollinator (Father)	OTT	Est. Weight	Pct.	Chart
1	1,506.00	Cardwell, Harvey		Oregon	United States	1790.5 Wallace	sib	384.0	1,266.00	19.0	
2	1,334.00	Sestak, Anne	Sublimity	Oregon	United States	2				-6.0	
3	1,230.50	Cohen, Ellann		Oregon	United States	1				-15.0	
EXH	1,145.00	Cardwell, Harvey		Oregon	United States	1	Warren	Oregon	45.8187717	-122.8501226	7.0
EXH	1,120.50	Ford, Iris		Oregon	United States	1	Glenoma	Washington	46.5145545	-122.1600994	16.0
4	1,117.50	Ford, Iris		Oregon	United States	2	Mulino	Oregon	45.2212415	-122.5816606	2.0
5	1,088.50	Pugh, Edson	Madras	Oregon	United States	1	Medford	Oregon	42.3230112	-122.8767396	3.0
6	1,074.50	Spielman, Kendall	Brooks	Oregon	United States	2				9.0	
7	1,031.50	Cowger, Tobias	RIDGEFIELD	Washington	United States	2				2.0	
Madras		Oregon				1	Sherwood	Oregon	45.3569341	-122.8410659	
Brooks		Oregon				1				-1.0	
RIDGEFIELD		Washington				1				-3.0	
Madras		Oregon				1	Portland	Oregon	45.515232	-122.6783853	3.0
Sublimity		Oregon				1	Lake Oswego	Oregon	45.4192504	-122.6629863	15.0
Dallas		Oregon				1	Warren	Oregon	45.8187717	-122.8501226	-9.0
Happy Valley		Oregon				1	Brooks	Oregon	45.048175	-122.9584291	13.0
						1	Dallas	Oregon	44.9192844	-123.3170472	5.0
						1	BONNEY LAKE	Washington	47.1770777	-122.1868601	-7.0
						1	Beaverton	Oregon	45.4869283	-122.804032	5.0
						1	Happy Valley	Oregon	45.4412329	-122.5125279	-10.0
						1	Dallas	Oregon	44.9192844	-123.3170472	9.0
EXH	515.00	Cardwell, Harvey				1	Sublimity	Oregon	44.8295663	-122.7945335	12.0
17	506.00	Seklreta, Matt	Medford	Oregon	United States	1	Madras	Oregon	44.6334544	-121.1294872	3.0
18	256.00	Sherwood, Jim	Mulino	Oregon	United States	2	RIDGEFIELD	Washington	47.6061389	-122.3328481	-6.0
19	161.00	Clayton, Brandon	Glenoma	Washington	United States	2	Brooks	Oregon	45.048175	-122.9584291	32.0
EXH	156.00	Possert, Michael	Warren	Oregon	United States	4	Madras	Oregon	44.6334544	-121.1294872	4.0
DMG	1,143.50	Frankfort, John		Oregon	United States					3.0	
DMG	806.50	Barker, Ron	BONNEY LAKE	Washington	United States	2501 Bernstrom	1997.5 Sherwood	334.0	855.00	-2.0	

28 Entries. (7 were exhibition only.)

Dirty Data



Converted to
add Longitude
& Latitude

Pumpkin Production



... did we make a better predictive model?

What did you try?

Lots of models!

- Linear regression
- Logistic regression
- Decision trees
- Random forest
- KNN (K nearest neighbor)
- PCA for feature selection



... but no pie



Our Updated Table

2025 GPC OTT Chart for Atlantic Giant Pumpkins

that you should be using instead

OTT	lbs	OTT	lbs	OTT	lbs	OTT	lbs	OTT	lbs	OTT	lbs	OTT	lbs	OTT	lbs	OTT	lbs	OTT	lbs	OTT	lbs
160	120.1	244	333.2	300	627.4	330	826.5	360	1064.7	390	1329.8	420	1626.1	450	1939.2	480	2181.0	510	2485.7		
165	120.1	246	344.8	301	649.3	331	826.5	361	1064.7	391	1329.8	421	1633.2	451	1986.9	481	2209.7	511	2485.7		
170	123.5	248	360.8	302	649.3	332	840.7	362	1095.6	392	1346.7	422	1626.9	452	1986.9	482	2251.0	512	2485.7		
175	137.5	250	369.0	303	650.5	333	843.8	363	1095.6	393	1355.0	423	1637.6	453	1986.9	483	2283.4	513	2485.7		
180	142.0	252	373.4	304	650.5	334	843.8	364	1102.4	394	1367.4	424	1637.6	454	1989.1	484	2283.4	514	2485.7		
185	154.6	254	386.7	305	672.6	335	878.0	365	1128.0	395	1367.4	425	1658.5	455	1989.1	485	2328.1	515	2485.7		
190	171.6	256	390.9	306	672.6	336	878.0	366	1128.0	396	1367.7	426	1676.8	456	2022.3	486	2371.1	516	2485.7		
195	181.1	258	428.3	307	672.6	337	878.0	367	1128.0	397	1396.7	427	1678.1	457	2022.3	487	2371.1	517	2485.7		
200	189.1	260	432.8	308	672.6	338	886.3	368	1128.0	398	1396.7	428	1675.5	458	2020.4	488	2371.1	518	2485.7		
202	189.1	262	437.5	309	672.6	339	887.2	369	1141.4	399	1417.4	429	1720.6	459	2043.0	489	2420.2	519	2485.7		
204	199.8	264	448.6	310	687.7	340	887.2	370	1141.4	400	1417.4	430	1716.2	460	2043.0	490	2485.7	520	2485.7		
206	199.8	266	459.9	311	702.1	341	919.2	371	1146.7	401	1418.9	431	1736.8	461	2050.9	491	2485.7	521	2485.7		
208	210.1	268	464.5	312	703.2	342	919.2	372	1181.6	402	1426.2	432	1742.3	462	2050.9	492	2485.7	522	2485.7		
210	217.3	270	482.3	313	703.2	343	934.8	373	1181.6	403	1426.2	433	1739.8	463	2118.7	493	2485.7	523	2485.7		
212	238.5	272	491.0	314	720.9	344	955.8	374	1182.3	404	1452.8	434	1793.7	464	2110.7	494	2485.7	524	2485.7		
214	251.8	274	495.2	315	720.9	345	955.8	375	1201.2	405	1454.9	435	1791.0	465	2149.0	495	2485.7	525	2485.7		
216	251.8	276	514.3	316	720.9	346	955.8	376	1202.7	406	1454.9	436	1791.0	466	2158.3	496	2485.7	526	2485.7		
218	251.1	278	514.3	317	747.0	347	970.6	377	1206.5	407	1487.3	437	1791.0	467	2196.9	497	2485.7	527	2485.7		
220	259.6	280	516.6	318	747.0	348	970.6	378	1206.5	408	1492.2	438	1838.8	468	2177.7	498	2485.7	528	2485.7		
222	262.0	282	527.7	319	747.8	349	970.6	379	1206.5	409	1492.2	439	1838.5	469	2173.5	499	2485.7	529	2485.7		
224	262.0	284	527.7	320	762.1	350	980.8	380	1241.3	410	1492.2	440	1836.8	470	2169.4	500	2485.7	530	2485.7		
226	262.9	286	559.7	321	762.1	351	999.6	381	1241.3	411	1535.7	441	1836.8	471	2150.8	501	2485.7	531	2485.7		
228	279.3	288	559.7	322	762.1	352	999.6	382	1241.3	412	1550.5	442	1836.8	472	2150.8	502	2485.7	532	2485.7		
230	279.3	290	567.2	323	763.6	353	1008.6	383	1237.2	413	1551.5	443	1836.8	473	2153.8	503	2485.7	533	2485.7		
232	282.9	292	598.8	324	791.9	354	1029.1	384	1268.3	414	1551.5	444	1852.7	474	2153.8	504	2485.7	534	2485.7		
234	301.0	294	598.8	325	791.9	355	1029.1	385	1268.3	415	1551.5	445	1895.0	475	2153.8	505	2485.7	535	2485.7		
236	307.3	296	617.0	326	791.9	356	1029.1	386	1278.7	416	1577.8	446	1895.0	476	2160.5	506	2485.7	536	2485.7		
238	307.3	297	623.7	327	791.9	357	1039.4	387	1309.1	417	1577.8	447	1897.8	477	2155.9	507	2485.7	537	2485.7		
240	309.5	298	627.4	328	791.9	358	1039.4	388	1309.1	418	1625.3	448	1901.9	478	2181.9	508	2485.7	538	2485.7		
242	327.6	299	627.4	329	826.5	359	1061.7	389	1329.8	419	1626.1	449	1907.6	479	2181.0	509	2485.7	539	2485.7		



Our Performance

Place	Grower Name	OTT	Weight (lbs)	Formula Est. Weight	Formula Error	Model Est Weight	Model Error	Model vs Formula
1	Cardwell, Harvey	384	1,506.00	1,266.00	240.00	1,268.26	237.74	Our model is better
2	Sestak, Anne	401	1,334.00	1,421.00	-87.00	1,418.94	-84.94	Our model is better
3	Cohen, Ellann	404	1,230.50	1,449.00	-218.50	1,452.84	-222.34	Formula Wins
EXH	Cardwell, Harvey	361	1,145.00	1,068.00	77.00	1,064.68	80.32	Formula Wins
DMG	Frankfort, John	372	1,143.50	1,161.00	-17.50	1,181.60	-38.10	Formula Wins
EXH	Ford, Iris	348	1,120.50	963	157.50	970.60	149.90	Our model is better
4	Ford, Iris	364	1,117.50	1,093.00	24.50	1,102.43	15.07	Our model is better
5	Pugh, Edson	360	1,088.50	1,060.00	28.50	1,064.68	23.82	Our model is better
6	Spielman, Kendall	351	1,074.50	986	88.50	999.58	74.92	Our model is better
7	Cowger, Tobias	354	1,031.50	1,011.00	20.50	1,029.18	2.32	Our model is better
EXH	Pugh, Edson	356	1,020.50	1,027.00	-6.50	1,029.18	-8.68	Formula Wins
EXH	Sestak, Annie	353	967.5	1,002.00	-34.50	1,008.56	-41.06	Formula Wins
8	Guy, Gregory	339.5	922.5	897	25.5	887.17	35.33	Formula Wins
9	Kristensen, Gary	322	885.5	769	116.5	762.10	123.40	Formula Wins
10	Wollman, Liesl	345	858	939	-81	955.82	-97.82	Formula Wins
DMG	Barker, Ron	334	806.5	855	-48.5	843.79	-37.29	Our model is better
11	Edwards, Ron	314	806	714	92	720.90	85.10	Our model is better
EXH	Spielman, Kendall	315	757.5	720	37.5	720.90	36.60	Our model is better
12	Possert, Jr, Michael	327	745	804	-59	791.91	-46.91	Our model is better
13	Williams, Brian	301	658	629	29	649.28	8.72	Our model is better
14	Williams, Luke	302	575.5	636	-60.5	649.28	-73.78	Formula Wins
15	Farrester, Josh	283	572	523	49	527.66	44.34	Our model is better
16	Bahns, Brad	272	520	464	56	491.03	28.97	Our model is better
EXH	Cardwell, Harvey	279	515	501	14	514.29	0.71	Our model is better
17	Seklreta, Matt	286	506	540	-34	559.74	-53.74	Formula Wins
18	Sherwood, Jim	203	256	194	62	120.10	135.90	Formula Wins
19	Clayton, Brandon	188	161	155	6		161.00	Formula Wins
EXH	Possert, Michael	186	156	151	5		156.00	Formula Wins



Source: BigPumpkins.com
Shown: 2024 Tualatin Results

Feature Selection & Model Details

Boosting Piecewise Model

OTT \leq 256

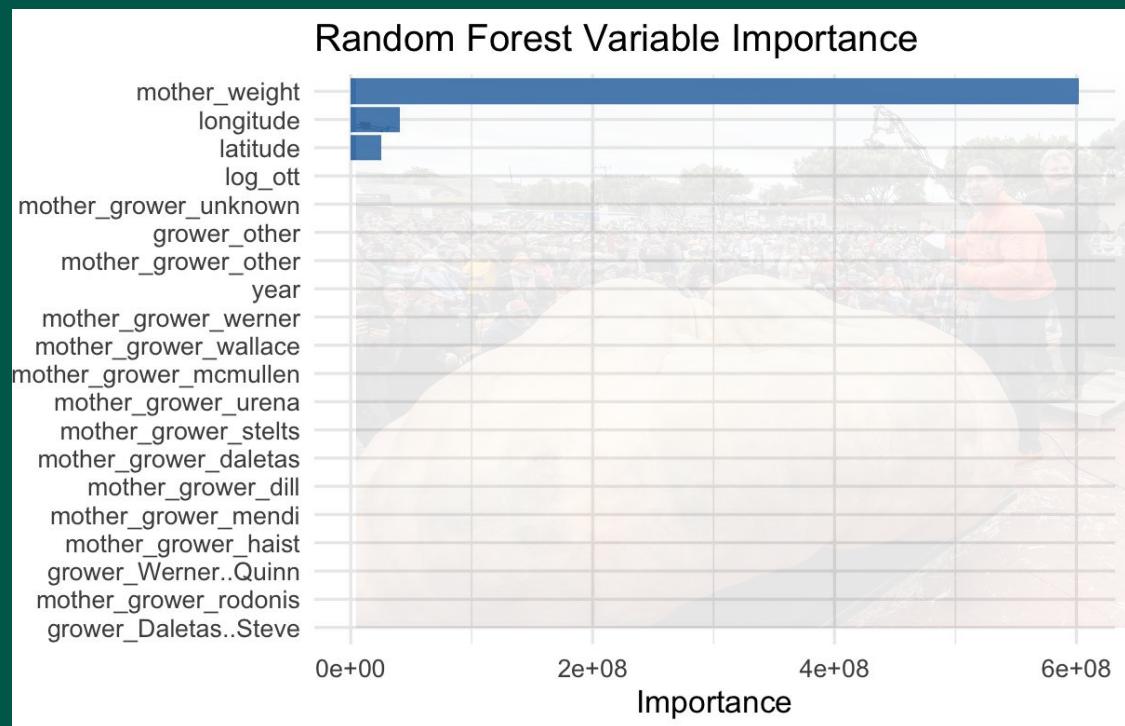
- Features: OTT only

OTT $>$ 256

- Features: OTT & Mother's weight

Hyperparameters autotuned

Model	RMSE OTT $>$ 256	RMSE OTT \leq 256
2025 Chart	99.4	26.0
Boosting	★ 87.7 ★	👑 26.6 🎉



Why were growers less important?

There were only ~9k unique growers
(across the 30k pumpkins)



Pumpkin Planning



... so what's next?

So what's next?

ShinyApp V1:

Predict current pumpkin's weight

- Includes OTT
- Mother Seed Origin Weight
- Father Pollinator Weight

ShinyApp V2:

Predict optimized seeds for a given location



Predicting Pumpkins Weights



Questions
or Comments?

5 pts: introduce your dataset to the audience.

- how many features? **OTT, 4 for parents, 2 for geo = 7** how many rows? **Approx 24k**
- continuous or categorical? missing data? **Both, ott is cont, parent is categorical, yes missing data**
- What is your dependent variable? **Weight**
- **regression** or classification? distribution?
- did you need to perform any data exploration, cleaning, transformation? **yes**

10 pts: perform some thoughtful supervised learning, including engineering and selecting features, selecting and optimizing a model, and explaining your model (coefficients or feature importance, performance).

- feature engineering / selection, bivariate charts? Interactions?
- missing data? how to handle it?
- Selection of modeling algorithm? classification or regression? binary or multi-class?
- interpretation of variable importance, coefficients if applicable
- justification of choice of metric (accuracy, precision / recall, other?)
- is class weighting or over / under sampling appropriate?
- discussion of choice or tuning of hyperparameters, if any
- meaningful discussion of predictive power and conclusions from model
- look at misclassified examples from test dataset, what do they say about your model?
- outliers in data? **Yes**

5 pts: Principal Component Analysis as data exploration and visualization.

- take a look at PCA, percent explained
- take a look at top eigenvector or two, what is it made out of?
- can you visualize your prediction problem by projecting to 2 dimensions?

5 pts: k-means as data exploration and visualization.

- discussion for choosing number of clusters
- analysis of cluster centers
- scatter plot(s) showing 2 dimensional perspective of clusters and cluster centers?
- meaningful interpretation / discussion of conclusions