

20141118	20141118_131037	5.109	N2	127.0	1	245.0	0.09477000000000001		0.675	0.887	0.243	1.127	0.36	1.171	0.863	0.0819	5.9	1.0
20141118	20141118_131037	5.149	N2	128.0	1	245.0	0.096228		0.6993	0.919	0.2403	1.115	0.344	1.118	0.9002	0.0814	7.5	1.0
20141118	20141118_131037	5.189	N2	129.0	1	245.0	0.096957		0.6993	0.919	0.2376	1.102	0.34	1.105	0.9039	0.0817	14.9	1.0
20141118	20141118_131037	5.228	N2	130.0	1	245.0	0.097686		0.7074	0.93	0.2349	1.09	0.332	1.08	0.9261	0.0845	6.3	1.0
20141118	20141118_131037	5.266	N2	131.0	1	245.0	0.095499		0.6993	0.919	0.2376	1.102	0.34	1.105	0.8585	0.0831	3.5	1.0
20141118	20141118_131037	5.308	N2	132.0	1	245.0	0.095499		0.7317	0.9620000000000001	0.2403	1.115	0.328	1.068	0.8855	0.0818	1.3	
20141118	20141118_131037	5.350	N2	133.0	1	245.0	0.095499	0.1633	14.7	0.756	0.993	0.243	1.127	0.321	1.046	0.9195	0.08	1.6
20141118	20141118_131037	5.390	N2	134.0	1	245.0	0.095499	0.1651	14.5	0.756	0.993	0.2484	1.152	0.329	1.069	0.8934	0.0797	4.0
20141118	20141118_131037	5.430	N2	135.0	1	245.0	0.096228	0.1669	14.4	0.756	0.993	0.2511	1.165	0.332	1.08	0.8707	0.0823	19.3
20141118	20141118_131037	5.468	N2	136.0	1	245.0	0.09477000000000001	0.1668	14.6	0.7533	0.99	0.2538	1.177	0.337	1.096	0.896	0.0809	
20141118	20141118_131037	5.511	N2	137.0	1	245.0	0.094041	0.1714	15.1	0.7587	0.997	0.2592	1.202	0.342	1.111	0.8945	0.0828	23.6
20141118	20141118_131037	5.551	N2	138.0	1	245.0	0.09477000000000001	0.1698	14.6	0.756	0.993	0.2619	1.215	0.346	1.127	0.8895	0.0827	
20141118	20141118_131037	5.591	N2	139.0	1	245.0	0.095499	0.1714	14.3	0.7533	0.99	0.2646	1.228	0.351	1.143	0.9129	0.0828	23.9
20141118	20141118_131037	5.631	N2	140.0	1	245.0	0.096957	0.1699	13.5	0.7533	0.99	0.2673	1.24	0.355	1.154	0.8685	0.0876	32.8
20141118	20141118_131037	5.670	N2	141.0	1	245.0	0.095499	0.1692	12.7	0.756	0.993	0.27	1.253	0.357	1.162	0.9007	0.085	22.6
20141118	20141118_131037	5.709	N2	142.0	1	245.0	0.097686	0.1647	11.9	0.7803	1.025	0.2754	1.278	0.353	1.148	0.9446	0.0843	33.4
20141118	20141118_131037	5.751	N2	143.0	1	245.0	0.09477000000000001	0.162	10.7	0.7776	1.022	0.2781	1.29	0.358	1.163	0.893	0.0825	
20141118	20141118_131037	5.792	N2	144.0	1	245.0	0.095499	0.1629	10.4	0.8046	1.057	0.2754	1.278	0.342	1.113	0.8582	0.0838	32.5
20141118	20141118_131037	5.831	N2	145.0	1	245.0	0.096957	0.1599	9.2	0.8073	1.061	0.2754	1.278	0.341	1.11	0.8949	0.0849	32.2
20141118	20141118_131037	5.870	N2	146.0	1	245.0	0.096957	0.1604	8.0	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949	0.0851	38.8
20141118	20141118_131037	5.908	N2	147.0	1	245.0	0.098415	0.1606	7.3	0.81	1.064	0.2862	1.328	0.353	1.149	0.9516	0.0833	34.4
20141118	20141118_131037	5.953	N2	148.0	1	245.0	0.096228	0.1634	5.7	0.7506	0.986	0.2727	1.265	0.363	1.182	0.911	0.0832	41.3
20141118	20141118_131037	5.992	N2	149.0	1	245.0	0.094041	0.1665	4.2	0.7479	0.983	0.2727	1.265	0.365	1.186	0.9179	0.0785	40.4
20141118	20141118_131037	6.032	N2	150.0	1	245.0	0.09185399999999999	0.1757	3.3	0.7452	0.9790000000000001	0.2808	1.303	0.377	1.226	0.9137		
20141118	20141118_131037	6.071	N2	151.0	1	245.0	0.092583	0.1769	2.0	0.7506	0.986	0.3078	1.428	0.41	1.334	0.8969	0.0773	45.1
20141118	20141118_131037	6.111	N2	152.0	1	245.0	0.096228	0.1834	1.1	0.7506	0.986	0.3078	1.428	0.41	1.334	0.9056	0.0808	44.3
20141118	20141118_131037	6.153	N2	153.0	1	245.0	0.094041	0.1876	2.5	0.7452	0.9790000000000001	0.2808	1.303	0.377	1.226	0.8954	0.0765	
20141118	20141118_131037	6.193	N2	154.0	1	245.0	0.093312	0.1919	3.7	0.7479	0.983	0.2808	1.303	0.375	1.221	0.9174	0.0765	43.6
20141118	20141118_131037	6.234	N2	155.0	1	245.0	0.09185399999999999	0.1974	4.6	0.7749	1.018	0.2754	1.278	0.355	1.156	0.897	0.0784	40.4
20141118	20141118_131037	6.273	N2	156.0	1	245.0	0.094041	0.2051	6.3	0.8019	1.054	0.2646	1.228	0.33	1.073	0.8949	0.0778	43.2
20141118	20141118_131037	6.335	N2	157.0	1	245.0	0.095499	0.2176	8.2	0.8019	1.054	0.2268	1.052	0.283	0.92	0.9049	0.0781	37.8
20141118	20141118_131037	6.376	N2	158.0	1	245.0	0.096228	0.2253	9.2	0.8046	1.057	0.2268	1.052	0.282	0.917	0.8655	0.0814	36.2
20141118	20141118_131037	6.415	N2	159.0	1	245.0	0.095499	0.2299	10.2	0.8262	1.0859999999999999	0.2241	1.04	0.271	0.882	0.8789	0.08	
20141118	20141118_131037	6.457	N2	160.0	1	245.0	0.094041	0.235	11.0	0.8019	1.054	0.2187	1.015	0.273	0.887	0.8833	0.0802	28.3
20141118	20141118_131037	6.496	N2	161.0	1	245.0	0.09477000000000001	0.2381	11.5	0.8235	1.082	0.216	1.002	0.262	0.853	0.8833	0.0809	
20141118	20141118_131037	6.534	N2	162.0	1	245.0	0.097686	0.2364	11.1	0.8532	1.121	0.2133	0.99	0.25	0.813	0.8994	0.0819	27.6
20141118	20141118_131037	6.574	N2	163.0	1	245.0	0.097686	0.2377	10.7	0.8208	1.079	0.2106	0.977	0.257	0.835	0.8583	0.0818	16.5
20141118	20141118_131037	6.613	N2	164.0	1	245.0	0.09914400000000001	0.2317	10.5	0.8424	1.107	0.2079	0.964	0.247	0.803	0.8699	0.0837	
20141118	20141118_131037	6.653	N2	165.0	1	245.0	0.096957	0.2297	10.8	0.8424	1.107	0.2052	0.952	0.244	0.792	0.9063	0.0823	28.3
20141118	20141118_131037	6.694	N2	166.0	1	245.0	0.095499	0.2307	11.2	0.8127	1.068	0.2025	0.939	0.249	0.81	0.8661	0.0809	8.8

# OUTLINE

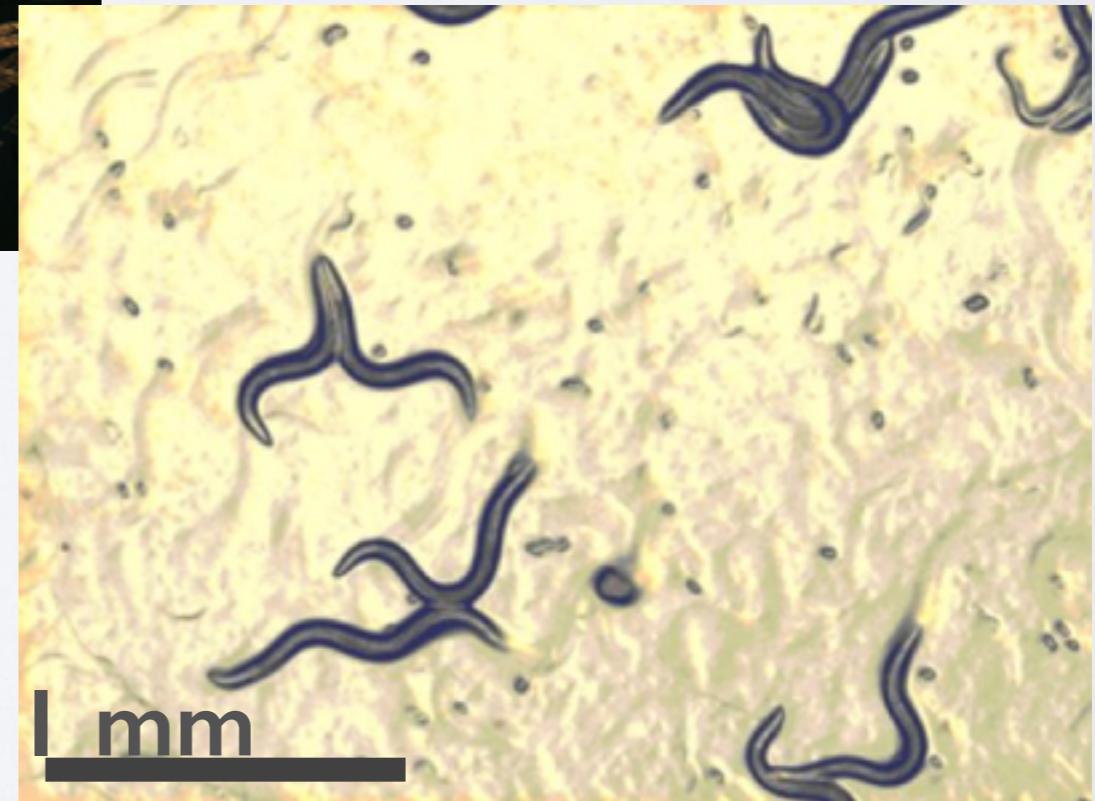
- A bit about my science
- My journey to using Python's Pandas
- Highlights of Pandas that makes this library super intuitive and welcoming for scientists

# A BIT ABOUT MY SCIENCE



Ph.D. in Neuroscience  
from 2005 - 2012

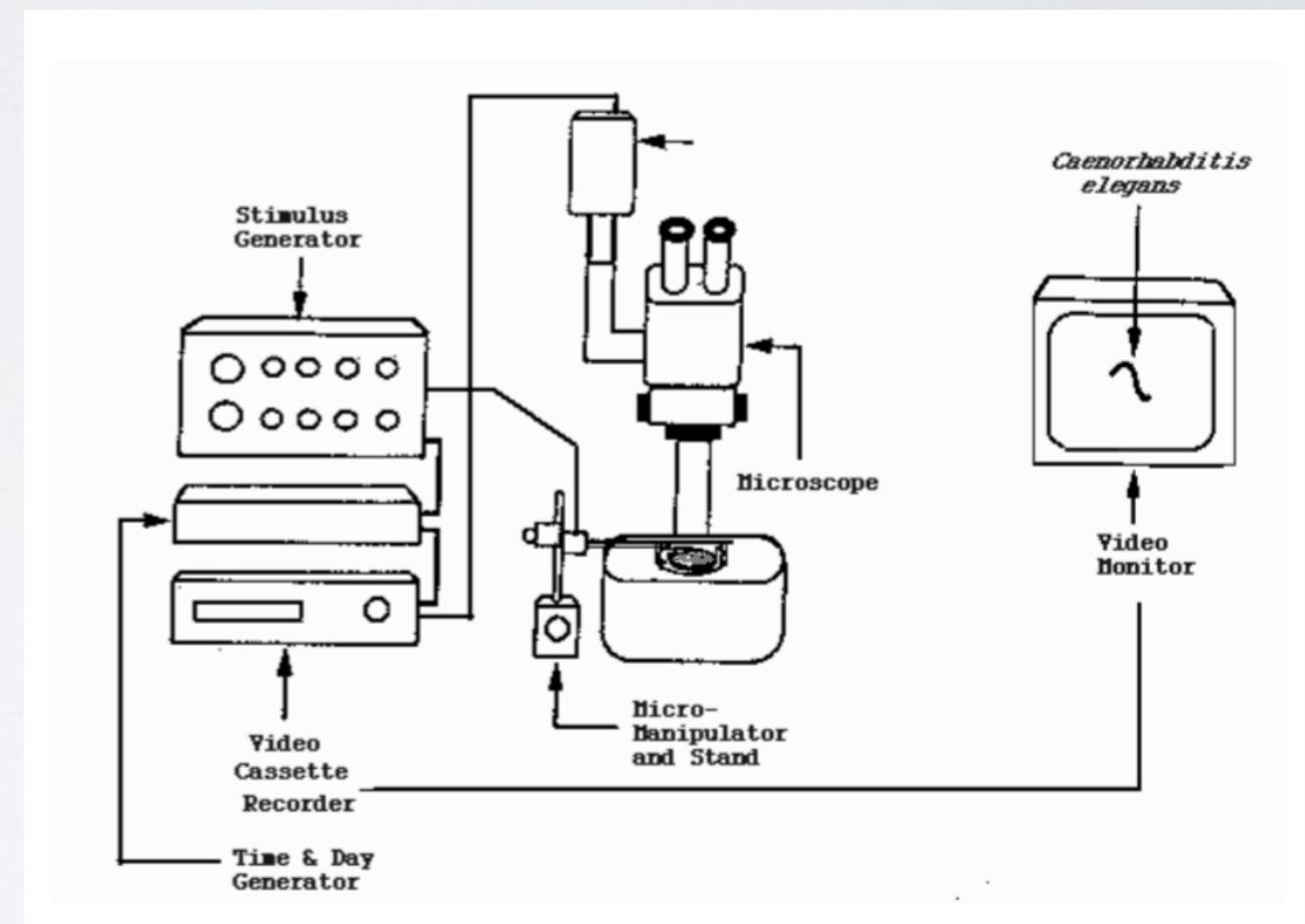
**Thesis:** Genetics of learning & memory in a microscopic nematode, *Caenorhabditis elegans*





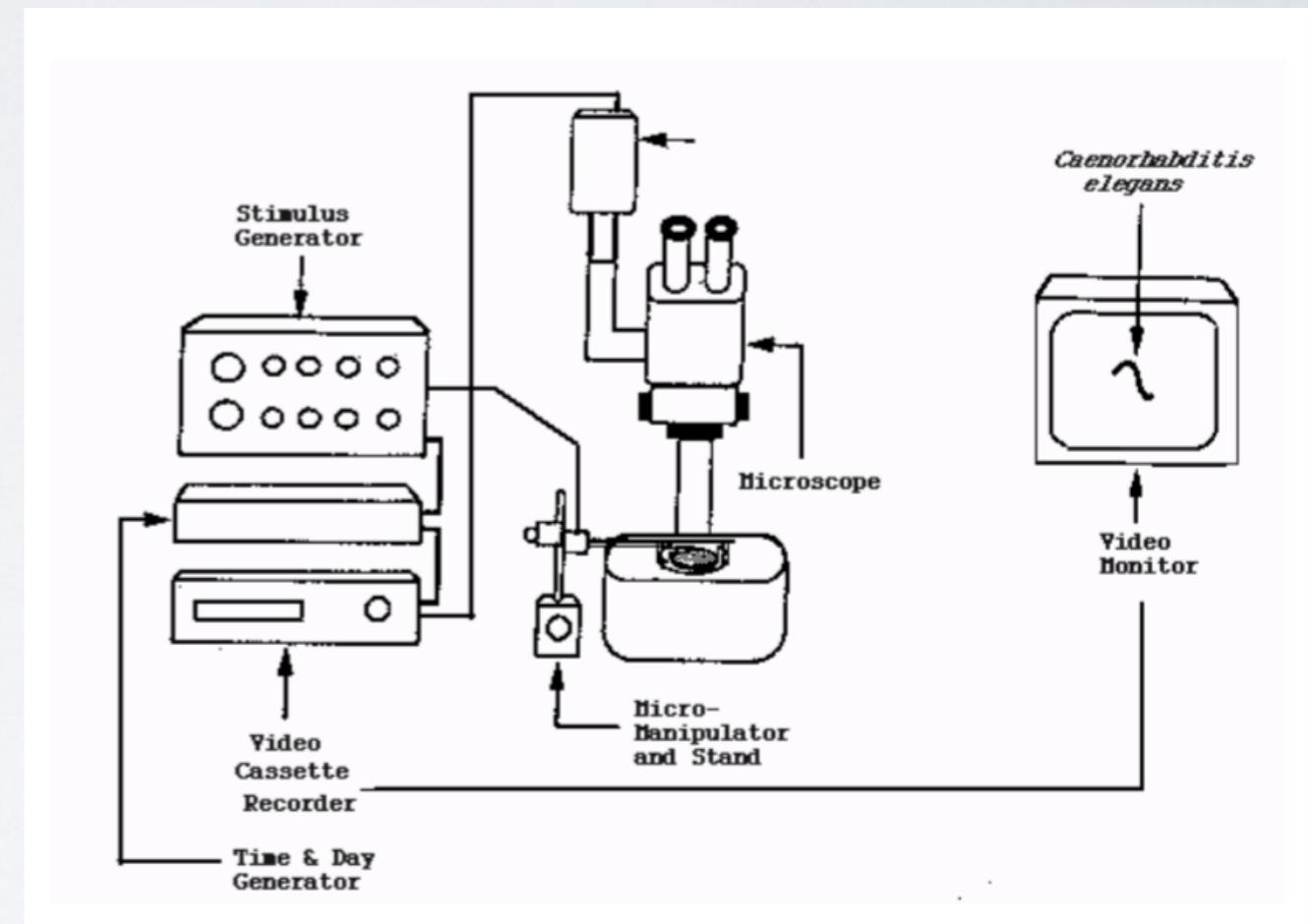
# 1990 - 2010: DATA COLLECTION BOTTLENECK

- Recorded a single worm at a time (5 - 30 min each)
- Re-watched videos to hand score
- Scanned into computer and measured with imageJ
- Used box-stats program to analyze and visualize data



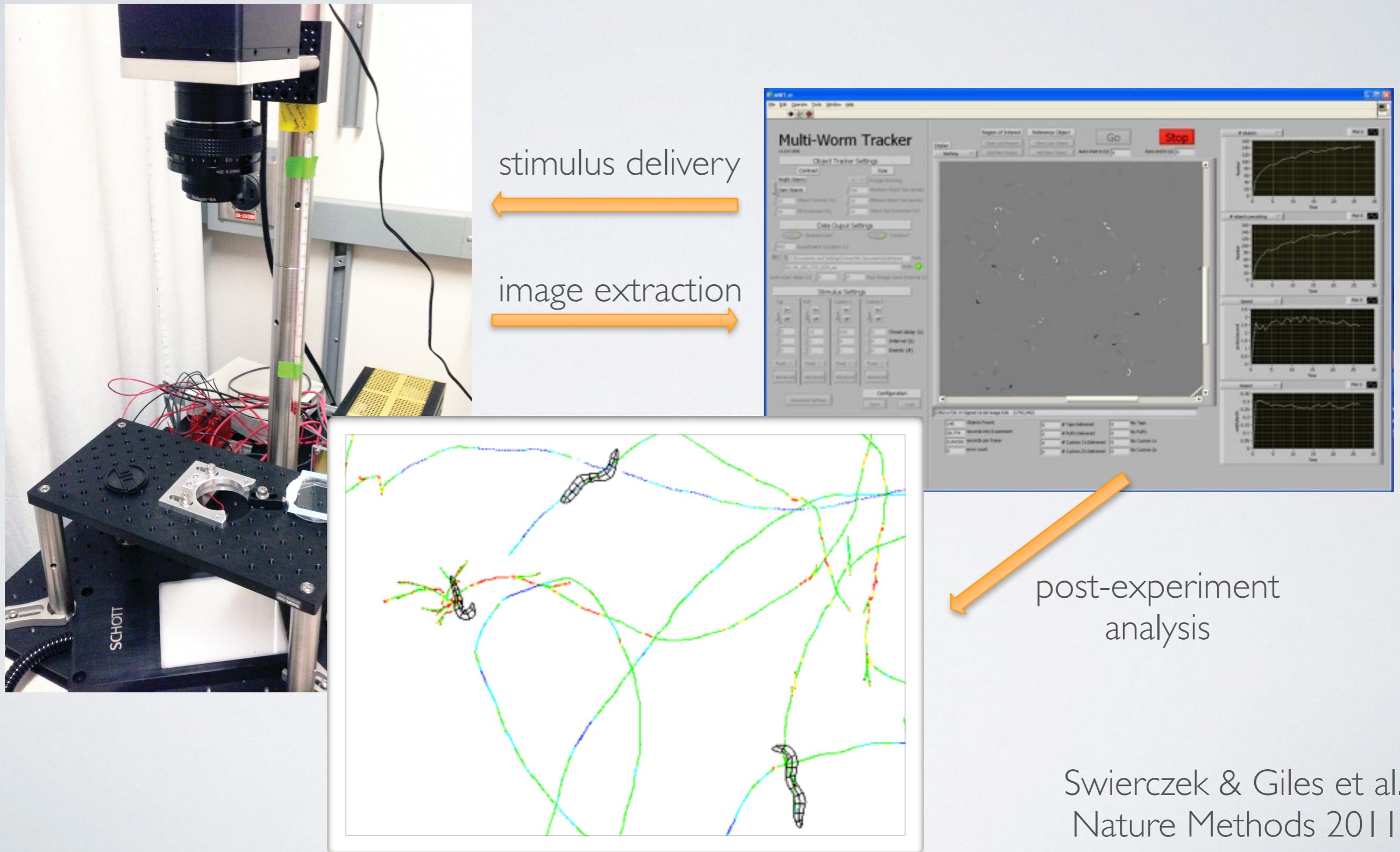
# 1990 - 2010: DATA COLLECTION BOTTLENECK

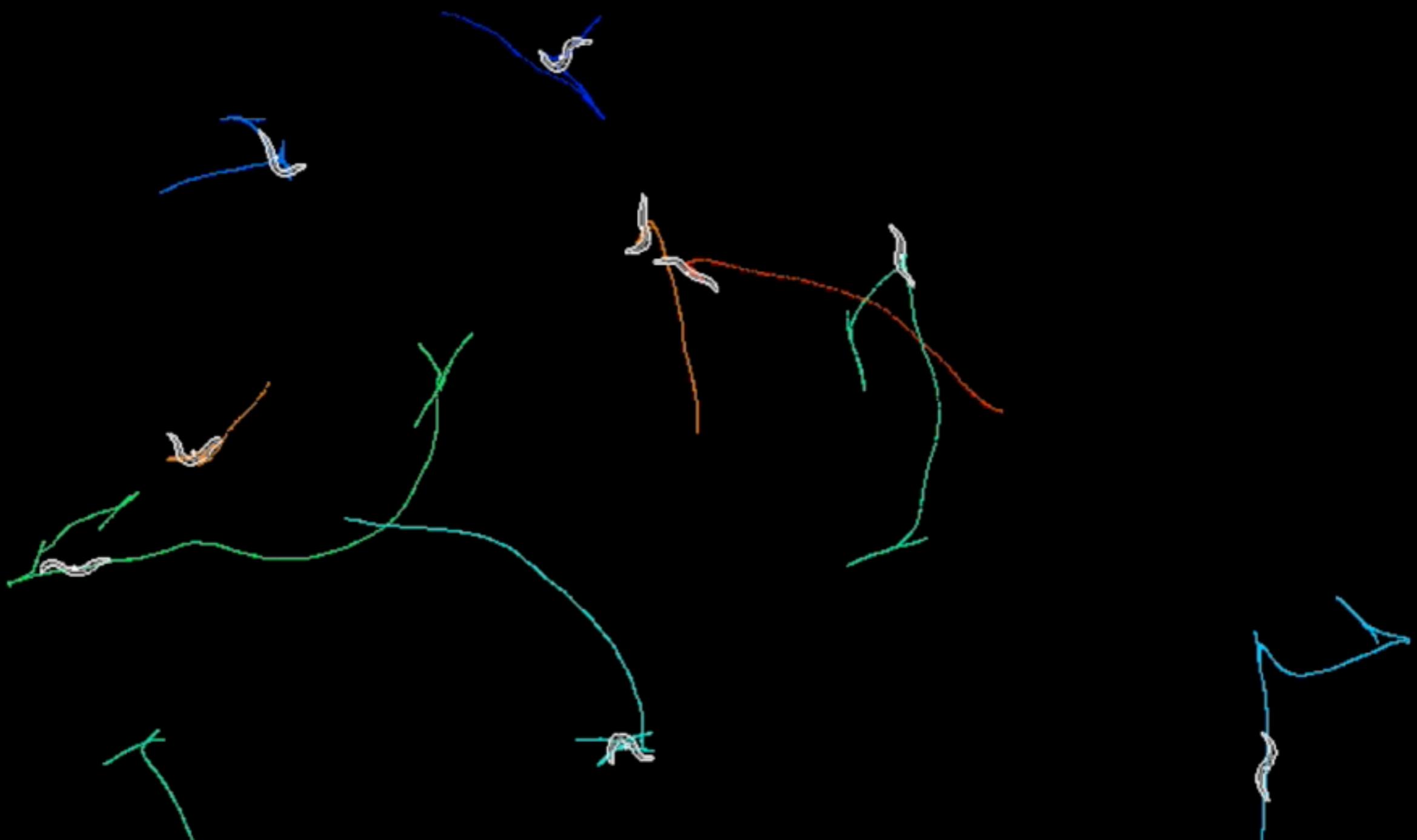
- Recorded a single worm at a time (5 - 30 min each)
- Re-watched videos to hand score
- Scanned into computer and measured with imageJ
- Used box-stats program to analyze and visualize data



*Very small datasets and time consuming!*

# THE MULTI-WORM TRACKER



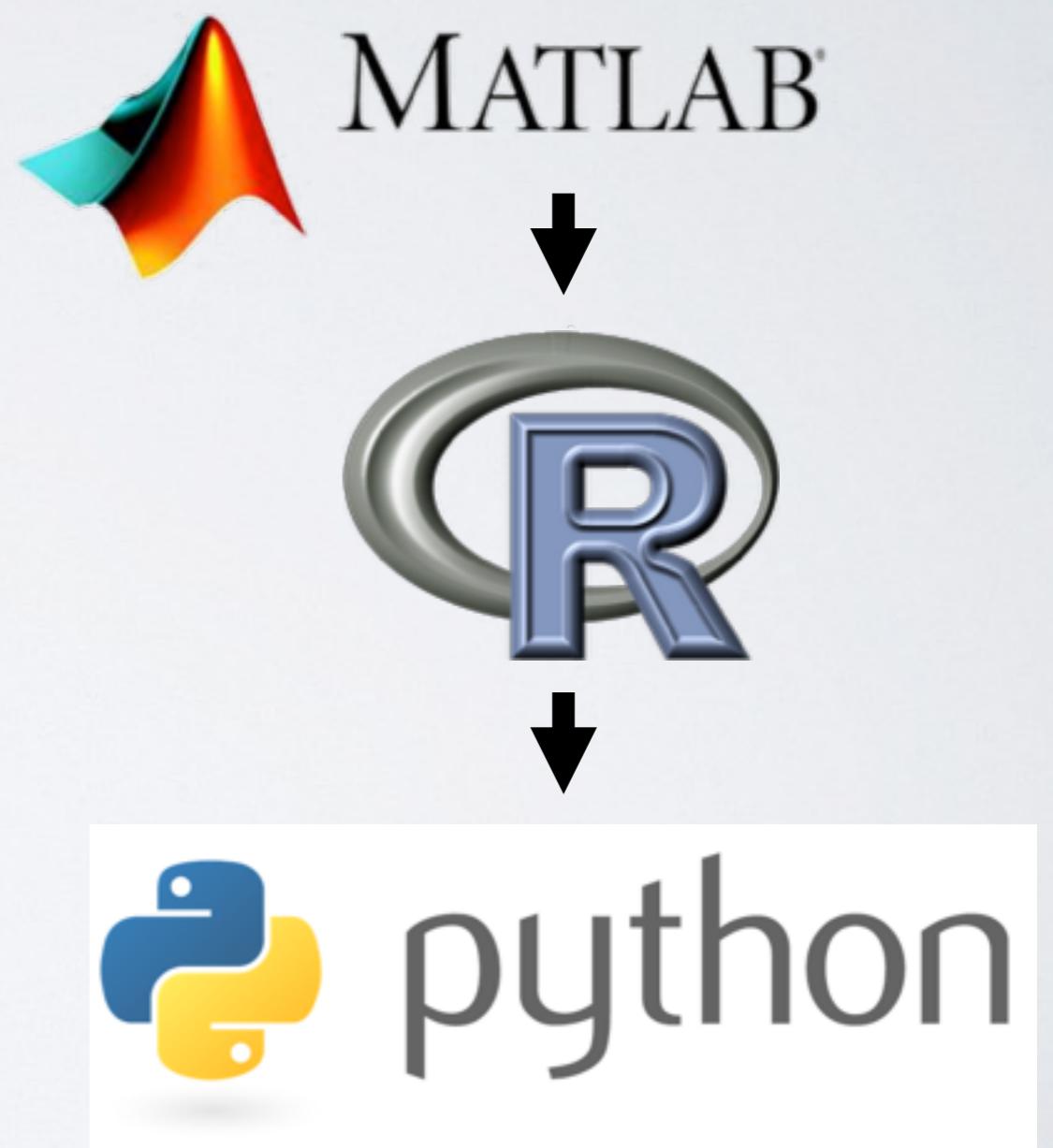


# 2010: DATA ANALYSIS BOTTLENECK

date	plate	time	strain	frame	id	persistance	area	speed	angular_speed	length	relalength	width	relwidth	aspect	relasp	
20141118	20141118_131037	5.065	N2	126.0	1	245.0	0.09477000000000001			0.675	0.887	0.243	1.127	0.36	1.171	0.8700
20141118	20141118_131037	5.109	N2	127.0	1	245.0	0.09477000000000001			0.675	0.887	0.243	1.127	0.36	1.171	0.863
20141118	20141118_131037	5.149	N2	128.0	1	245.0	0.096228		0.6993	0.919	0.2403	1.115	0.344	1.118	0.9002	0.0814
20141118	20141118_131037	5.189	N2	129.0	1	245.0	0.096957		0.6993	0.919	0.2376	1.102	0.34	1.105	0.9039	0.0817
20141118	20141118_131037	5.228	N2	130.0	1	245.0	0.097686		0.7074	0.93	0.2349	1.09	0.332	1.08	0.9261	0.0845
20141118	20141118_131037	5.266	N2	131.0	1	245.0	0.095499		0.6993	0.919	0.2376	1.102	0.34	1.105	0.8585	0.0831
20141118	20141118_131037	5.308	N2	132.0	1	245.0	0.095499		0.7317	0.9620000000000001	0.2403	1.115	0.328	1.068	0.0831	
20141118	20141118_131037	5.350	N2	133.0	1	245.0	0.095499	0.1633	14.7	0.756	0.993	0.243	1.127	0.321	1.046	0.9195
20141118	20141118_131037	5.390	N2	134.0	1	245.0	0.095499	0.1651	14.5	0.756	0.993	0.2484	1.152	0.329	1.069	0.8934
20141118	20141118_131037	5.430	N2	135.0	1	245.0	0.096228	0.1669	14.4	0.756	0.993	0.2511	1.165	0.332	1.08	0.8707
20141118	20141118_131037	5.468	N2	136.0	1	245.0	0.09477000000000001	0.1668	14.6	0.7533	0.99	0.2538	1.177	0.337	1.096	0.9129
20141118	20141118_131037	5.511	N2	137.0	1	245.0	0.094041	0.1714	15.1	0.7587	0.997	0.2592	1.202	0.342	1.111	0.8945
20141118	20141118_131037	5.551	N2	138.0	1	245.0	0.09477000000000001	0.1698	14.6	0.756	0.993	0.2619	1.215	0.346	1.127	0.9129
20141118	20141118_131037	5.591	N2	139.0	1	245.0	0.095499	0.1714	14.3	0.7533	0.99	0.2646	1.228	0.351	1.143	0.9129
20141118	20141118_131037	5.631	N2	140.0	1	245.0	0.096957	0.1699	13.5	0.7533	0.99	0.2673	1.24	0.355	1.154	0.8685
20141118	20141118_131037	5.670	N2	141.0	1	245.0	0.095499	0.1692	12.7	0.756	0.993	0.27	1.253	0.357	1.162	0.9007
20141118	20141118_131037	5.709	N2	142.0	1	245.0	0.097686	0.1647	11.9	0.7803	1.025	0.2754	1.278	0.353	1.148	0.9446
20141118	20141118_131037	5.751	N2	143.0	1	245.0	0.09477000000000001	0.162	10.7	0.7776	1.022	0.2781	1.29	0.358	1.163	0.9446
20141118	20141118_131037	5.792	N2	144.0	1	245.0	0.095499	0.1629	10.4	0.8046	1.057	0.2754	1.278	0.342	1.113	0.8582
20141118	20141118_131037	5.831	N2	145.0	1	245.0	0.096957	0.1599	9.2	0.8073	1.061	0.2754	1.278	0.341	1.11	0.8949
20141118	20141118_131037	5.870	N2	146.0	1	245.0	0.096957	0.1604	8.0	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	5.909	N2	147.0	1	245.0	0.096957	0.1604	7.7	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	5.947	N2	148.0	1	245.0	0.096957	0.1604	7.4	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	5.986	N2	149.0	1	245.0	0.096957	0.1604	7.1	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.024	N2	150.0	1	245.0	0.096957	0.1604	6.8	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.063	N2	151.0	1	245.0	0.096957	0.1604	6.5	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.101	N2	152.0	1	245.0	0.096957	0.1604	6.2	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.139	N2	153.0	1	245.0	0.096957	0.1604	5.9	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.178	N2	154.0	1	245.0	0.096957	0.1604	5.6	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.216	N2	155.0	1	245.0	0.096957	0.1604	5.3	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.254	N2	156.0	1	245.0	0.096957	0.1604	5.0	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.293	N2	157.0	1	245.0	0.096957	0.1604	4.7	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.331	N2	158.0	1	245.0	0.096957	0.1604	4.4	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.369	N2	159.0	1	245.0	0.096957	0.1604	4.1	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.407	N2	160.0	1	245.0	0.096957	0.1604	3.8	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.445	N2	161.0	1	245.0	0.096957	0.1604	3.5	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.483	N2	162.0	1	245.0	0.096957	0.1604	3.2	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.521	N2	163.0	1	245.0	0.096957	0.1604	2.9	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.559	N2	164.0	1	245.0	0.096957	0.1604	2.6	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.597	N2	165.0	1	245.0	0.096957	0.1604	2.3	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.635	N2	166.0	1	245.0	0.096957	0.1604	2.0	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.673	N2	167.0	1	245.0	0.096957	0.1604	1.7	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.711	N2	168.0	1	245.0	0.096957	0.1604	1.4	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_131037	6.749	N2	169.0	1	245.0	0.096957	0.1604	1.1	0.8046	1.057	0.2781	1.29	0.346	1.124	0.8949
20141118	20141118_1															

# SOLUTION = PROGRAMMING LANGUAGES

- Initially started with Matlab
- Moved to R for access to data frames and advanced statistic capabilities
- Still work in R a lot, but moving into Python because of culture and ease to learn and readability of code



# PANDAS: A BRIEF HISTORY

- First released to open source by Wes McKinney (AQR) in 2009
- Wanted R statistical capabilities and better data-munging abilities in a more readable language
- Inspired by Jonathon Taylor's (Stanford) port of R's MASS package to Python

# PANDAS TODAY:

Pandas is a library that makes analysis of complex tabular data **easy!**

- ***Data frame data structures***: database-like tables with rows and columns
- ***Easy time series manipulations***
- ***Quick visualizations*** based on Matplotlib
- Very ***flexible import and export*** of data
- ***Data munging***: remove duplicates, manage missing values, automatically join tables by index
- ***SQL-like operations***: join, aggregate (group by)