

Introduction

The corpus used for this project consisted of 19 Greek and Roman (17 Greek, 2 Roman) works with some mythological content. These can broadly be grouped into poems/epics or plays (specifically, tragedies). The plays come from Aeschylus, Euripides, and Sophocles, and the other works are by Homer, Hesiod, Ovid, and Vergil; additionally, the *Homeric Hymns* are included, and the author is unknown for these.

Some of the discussion in this report is repeated within the Jupyter notebook, but I have tried to focus more on results here and more on process within the notebook.

Text Parsing

Most of the works we used as examples in classwork followed a novel-like format, so they had a somewhat standard hierarchy of content: a book is typically outlined by chapters; each chapter is outlined by paragraphs; each paragraph is outlined by sentences; each sentence is made up of words.

With these poems and plays, the hierarchy was not as clear to me. For more narratively driven poems, the structure can be similar to a novel, but individual lines can follow a meter that further adds structure and meaning to words. For dialogue-driven plays, it is not as clear what should be a “paragraph” or a “chapter”, especially since these plays do not have markings like “Acts” or “Scenes” you might find in a modern work. These works were additionally not originally written in English, so there is undoubtedly some meaning which is lost in translation.

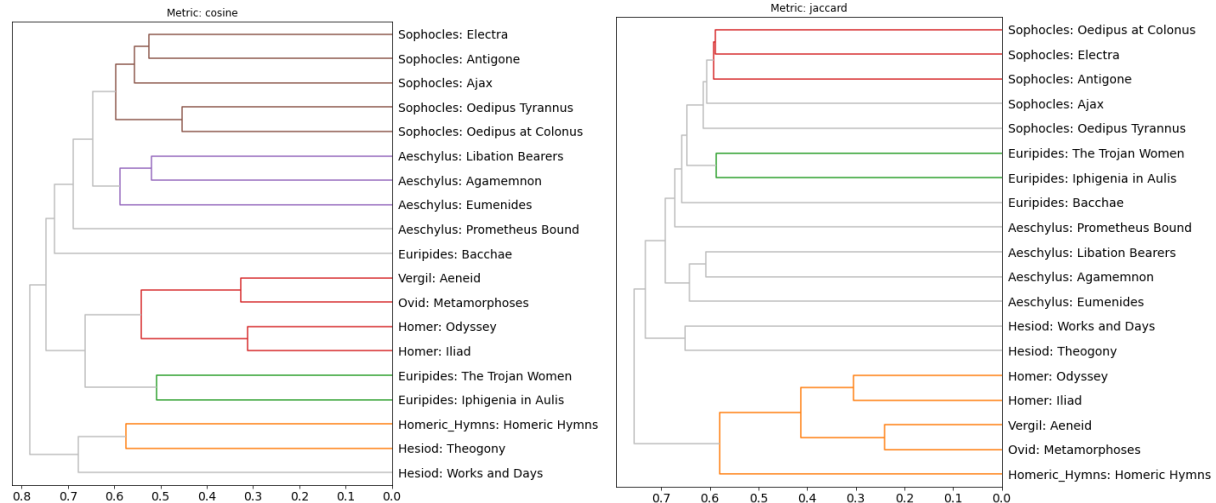
Overall, I broke up works based on chunking that the Perseus Digital Library uses in displaying each text, and for plays, specifically, I considered “paragraphs” to be changes in speakers. I think it has produced interesting insights, but it is important to consider the ways in which this approach could be lacking and improved.

Hierarchical Clustering

I started my analysis by taking the top 6000 terms according to DF-IDF and computing the pairwise cosine and Jaccard distances of each term’s maximum-normalized TF-IDF vectors (Jaccard uses binary vectors, however – term is present/not present). Dendrograms were created for each of these metrics using complete linkage. These trees are shown below.

In both trees, an author’s works tend to cluster together. One work by Aeschylus, *Prometheus Bound*, is somewhat separated from the other three works of Aeschylus included here; apparently, some scholars contest that Aeschylus wrote *Prometheus Bound*, so that could be interesting to look further into. However, the other three works comprise a trilogy, so it could just be that content overlaps more between these three.

The two epics of Homer (*Iliad* and *Odyssey*) and the two Roman works (*Aeneid* and *Metamorphoses*) also cluster together. The *Aeneid*, by Vergil, is thought to be heavily influenced by Homer’s works. *Metamorphoses* is more like a collection of mythological stories, but it does cover some of the same material as the other epic works and may have a similar style.



Hesiod's *Theogony* and the *Homeric Hymns* are quite close under cosine distance. *Theogony* describes the origins of Greek gods, and the *Homeric Hymns* are a collection of poems celebrating individual Greek gods. The explicit focus on gods could explain their similarity.

Principal Components Analysis

I performed PCA on the maximum-normalized TF-IDF vectors of the top 6000 terms, filtered as before using DF-IDF. Scatterplots and distributional plots of PC1 vs. PC0 and PC2 vs. PC1 are displayed below. The plots in the notebook are interactive, so please see those to explore the visualizations more easily.



PC0 (the first principal component which explains the highest proportion of variance in the TF-IDF vectors) seems to primarily be distinguishing works based on whether a work is a play. An exception is that the *Homeric Hymns* cluster with the plays in the PC0 dimension.

PC1 can divide the works clearly between Greek and Roman works, and more broadly, it may indicate evolution of style over time. On one end of PC1, there are the oldest works: the *Homeric Hymns* and Homer's epics. Towards the middle, there are the plays by Euripides, Sophocles, and Aeschylus. On the other end of PC1 are the two Roman authors Vergil and Ovid, who are the most "modern" in this corpus. There is an exception here as well: Hesiod's works fall among the plays.

It is not entirely clear aspect of the texts PC2 is modeling, if anything, and I did not examine principal components any further than that.



Topic Models (LDA)

A Latent Dirichlet Allocation (LDA) topic model was fit to the corpus. When applied to a document-term matrix (of TF-IDF values, in this case), this produces a document-topic table and a topic-word table. The documents used for this model were chapter-like chunks of each work. Additionally, each topic is represented by the top 10 words associated with it. 40 topics were generated to represent the corpus (thus, there are 400 total words representing the topics).

Among the 400 words representing topics, highly occurring words were "man" (19), "gods" (18), "men" (15), "land" (14), and "house" (14). Certainly, I think it's fair to say much of Greek mythology focuses on heroic men and the gods, who both bless mortals and come into conflict with them for various reasons.

A bar chart of topic weights along with the top 10 words that represent each topic is available in the notebook; it shows an ordered list of how associated certain topics are with the entire corpus.

What I find particularly interesting, however, is the table of authors and topics, shown below. Topics are ordered in numerical order, and darker cells in an author's column indicate a topic that author is more strongly associated with. The top 10 words associated with a topic are shown in the 'label' column.

	author	Aeschylus	Euripides	Hesiod	Homer	Homeric_Hymns	Ovid	Sophocles	Vergil	label
topic_id										
0	0.024961	0.060186	0.000013	0.000019		0.003460	0.000014	0.013372	0.000012	0 son, ships, earth, mountain, land, horses, course, offspring, justice, gods
1	0.000326	0.000192	0.000013	0.000019		0.059661	0.000014	0.013328	0.000012	1 men, goddess, life, power, chamber, crafts, success, prosperity, daughter goddess, city
2	0.080109	0.000192	0.000013	0.000019		0.001119	0.000014	0.013403	0.000012	2 house, man, stranger, heart, news, message, grief, gods, friends, way
3	0.003261	0.011123	0.000013	0.000809		0.001119	0.020192	0.007000	0.468083	3 war, hand, arms, gods, sword, way, foe, thee, power, death
4	0.000326	0.030107	0.000013	0.000019		0.030445	0.000014	0.000318	0.000012	4 mother, woe, child, day, trouble, father, bitter, messenger, man, bringing
5	0.001741	0.006428	0.000013	0.110083		0.012085	0.003280	0.013400	0.006165	5 ship, comrades, men, sea, thou, man, land, heart, ships, way
6	0.001147	0.001078	0.000013	0.021948		0.022481	0.001098	0.000318	0.000864	6 ships, son, sons, host, men, counsel, king, war, hath, people
7	0.000326	0.000192	0.000013	0.000019		0.001119	0.000014	0.000318	0.000012	7 abide, rebuke, refrain, reef, reeds, reed, red, recompense, reckless, recess
8	0.317842	0.199719	0.068053	0.003439		0.002440	0.099289	0.439889	0.073289	8 man, father, gods, land, city, mother, house, death, hands, life
9	0.000326	0.059973	0.000013	0.000019		0.006061	0.000332	0.013288	0.000012	9 gods, son, day, channels, horses, land, sea, tresses, blooms, shields
10	0.040567	0.000192	0.000013	0.000019		0.001119	0.000014	0.000318	0.000012	10 hand, death, chorus, father, man, men, justice, speech, home, enemy
11	0.000326	0.021908	0.000013	0.020653		0.016653	0.000014	0.007609	0.000012	11 wall, ships, men, son, war, twain, battle, horses, man, ground
12	0.013316	0.002676	0.050224	0.000813		0.026601	0.502979	0.019503	0.024505	12 love, son, death, time, life, blood, waves, hand, words, father
13	0.029430	0.041423	0.000013	0.000019		0.028801	0.000014	0.063624	0.000012	13 father, child, gods, man, daughter, death, help, life, hand, words
14	0.000326	0.032862	0.000013	0.000019		0.062611	0.000014	0.066458	0.000012	14 man, men, city, gods, land, mother, place, way, head, seat
15	0.061941	0.043992	0.000013	0.000019		0.001119	0.000014	0.013492	0.000012	15 woe, house, life, altar, child, goddess, ah, heart, head, army
16	0.001189	0.032772	0.000013	0.000315		0.008495	0.000014	0.028251	0.000012	16 land, city, children, god, things, man, men, people, whichever, gods
17	0.041456	0.025406	0.000013	0.000019		0.021253	0.333044	0.018248	0.304552	17 sea, love, land, arms, eyes, waves, words, way, voice, heart
18	0.000326	0.000192	0.000013	0.000019		0.001119	0.000014	0.000318	0.000012	18 abide, rebuke, refrain, reef, reeds, reed, red, recompense, reckless, recess
19	0.000481	0.000192	0.000013	0.028435		0.009551	0.000223	0.000822	0.000012	19 bow, wooers, house, hall, heart, man, son, hands, men, arrow
20	0.020494	0.026554	0.000013	0.000019		0.001119	0.000014	0.000318	0.000012	20 house, fortune, wealth, prostrate ground, rise, misery, triumph, prostrate, escape, shout
21	0.000756	0.000192	0.000013	0.000019		0.002414	0.000014	0.028067	0.000012	21 ah, ah ah, hope, race, chorus, house, power, life mortals, mortals, ruin
22	0.011777	0.006012	0.157401	0.353482		0.015808	0.017677	0.013288	0.051761	22 thou, man, son, men, thee, heart, house, gods, land, wooers
23	0.007558	0.047441	0.000013	0.000019		0.001119	0.000014	0.000753	0.000012	23 man, daughter, child, army, husband, letter, marriage, matter, goddess, hand
24	0.022108	0.033591	0.000013	0.000019		0.001119	0.000377	0.049884	0.000012	24 child, death, corpse, mother, life, house, son, fate, man, wife
25	0.000326	0.000192	0.000013	0.000019		0.001119	0.000014	0.026257	0.000012	25 life, father, house, voice, boy, word, tomb, son, grief, land
26	0.138606	0.032334	0.000013	0.000019		0.001119	0.000014	0.007002	0.000012	26 gods, man, house, heart, men, home, wealth, time, hand, mortals
27	0.003412	0.030130	0.000013	0.000019		0.001119	0.000014	0.030235	0.000012	27 lock, gods, land, place, city, son, ship, awaits, hand, turn
28	0.044257	0.000192	0.000013	0.000019		0.029258	0.000014	0.015039	0.000012	28 man, words, father, god, land, things, theseus, arts, children, dearest
29	0.010129	0.109866	0.000013	0.000019		0.041990	0.000076	0.006464	0.000012	29 house, men, father, gods, god, force, hands, city, land, son
30	0.030035	0.000192	0.000013	0.026796		0.001119	0.000062	0.000606	0.004256	30 heart, land, house, thou, men, thee, raft, halls, gods, goddess
31	0.000326	0.022141	0.000013	0.000019		0.001119	0.000014	0.005436	0.017638	31 shield, way, words, spear, slain, sword, troop, life, sea, stood
32	0.026887	0.026952	0.723836	0.035122		0.511091	0.016808	0.028716	0.012677	32 gods, men, son, earth, sea, heart, goddess, daughter, deathless, god
33	0.021016	0.001780	0.000013	0.022493		0.001119	0.000277	0.017698	0.000012	33 thou, land, city, father, thee, man, maidens, heart, mother, gods
34	0.000326	0.087937	0.000013	0.003964		0.001119	0.000014	0.016354	0.000012	34 god, hands, man, house, women, rites, master, thyrsos, things, city
35	0.020538	0.000192	0.000013	0.000019		0.002539	0.000014	0.000318	0.000012	35 bidding, gods, tales, things, sword, brand, lord, passions, arms, deeds
36	0.000579	0.000192	0.000013	0.000019		0.030804	0.000014	0.000392	0.000012	36 gold, gods hands, song, gods, sea, waves sea, set, hands, neck, home
37	0.020494	0.000192	0.000013	0.000019		0.001119	0.000014	0.000318	0.000012	37 house, heart, cloth, grief, blood, dreams, mother, life, sound, libations
38	0.000326	0.000192	0.000013	0.000019		0.001119	0.000014	0.013798	0.000012	38 man, hand, men, goal, gods, time, arms, eyes, flocks, sword
39	0.000326	0.002914	0.000013	0.371155		0.035352	0.003942	0.005790	0.035866	39 son, thou, man, spear, ships, men, battle, heart, horses, thee

All three tragedy playwrights are associated most highly with topic 8, represented by “man, father, gods, land, city, mother, house, death, hands, life”.

Hesiod and the *Homeric Hymns* focus a lot on the gods themselves, rather than gods in interaction with humans, and they both are most associated with topic 32, represented by “gods, men, son, earth, sea, heart, goddess, daughter, deathless, god”.

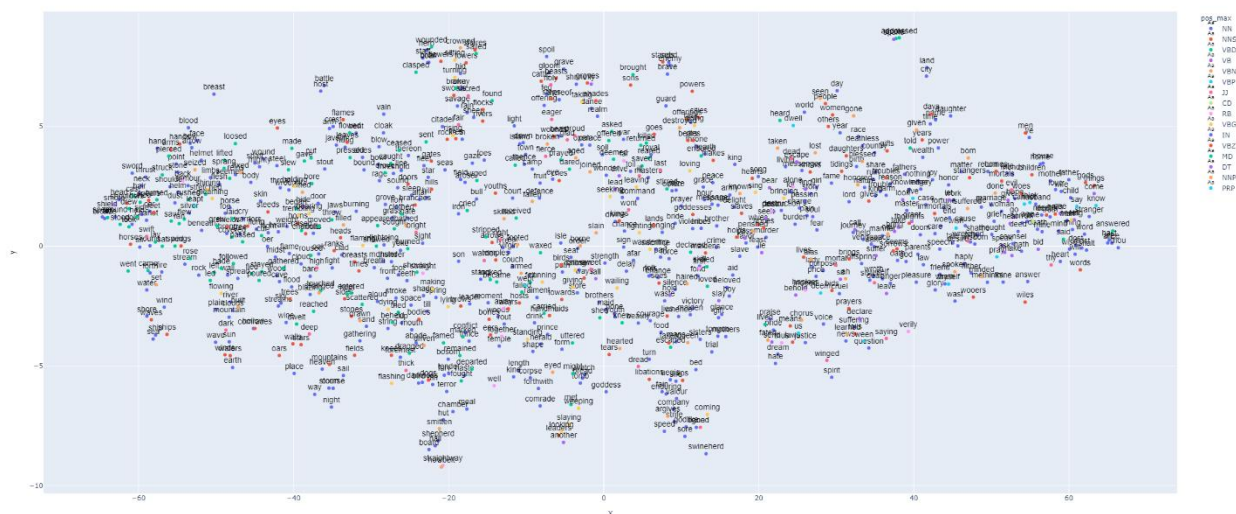
Vergil's singular work in this corpus, the *Aeneid*, can be broken into two main sections: the journey from Troy to Italy after the end of the Trojan War, and the war they fight in Italy

after their arrival. Correspondingly, Vergil is highly associated with two topics. Topic 17 is represented by “sea, love, land, arms, eyes, waves, words, way, voice, heart”; this could represent the journey section of the *Aeneid*. Topic 3 is represented by “war, hand, arms, gods, sword, way, foe, thee, power, death”; this could represent the war section of the *Aeneid*.

Overall, I feel like the topic model does a good job of picking out themes in the corpus that line up reasonably well with the actual works.

Word Embeddings (word2vec)

Word embeddings create dense representations of words based on statistics of co-occurrence they have with other words. Words that appear in similar contexts should have similar embedding vectors. The word2vec algorithm generated vectors with 256 dimensions, but a 2-dimensional representation of these vectors (using t-SNE) is shown below; an interactive version is available in the notebook which is much easier to explore.



In this plot, some clusters of related words stuck out to me. There are descriptions of the world, particularly the ocean, that are close together: “sea”, “ship”, “shore”, “waves”, “water”, “wind”, “earth”, “sun”, “dark”. There are also words that are used for descriptions of people, their injuries, and their armaments which are close together: “head”, “shoulders”, “hair”, “neck”, “hand”, “blood”, “bronze”, “spear”, “shield”. Lastly, several descriptions of emotions and concepts are close together: “justice”, “suffering”, “wroth”, “vengeance”, “sorrows”, “pleasure”, “glory”, “desire”. Certainly, other clusters of related words could be picked out with additional exploration.

It is also possible to perform “analogies” with word vectors by adding and subtracting them. The analogy “A : B :: C : ?” can be emulated with the vector algebra “A – B + C = ?”. Here are some examples I performed.

- War – bronze + peace = gold
- Sea – death + earth = life
- Sword – shield + chariot = horses

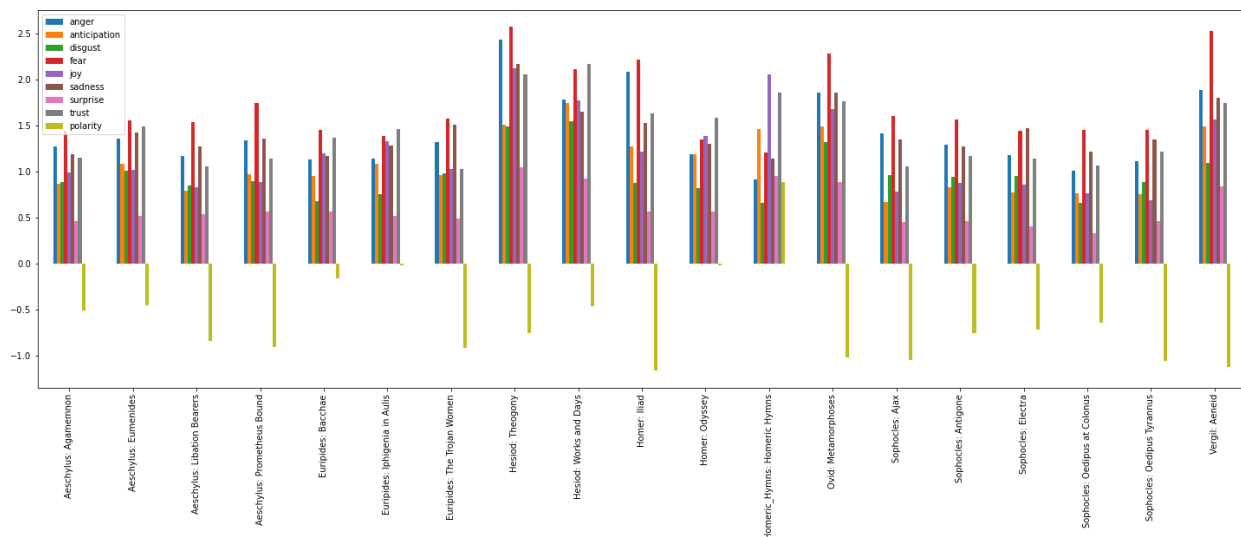
These might be interpreted in various ways, and they do not always result in something that makes sense to a human, but these work well, I think. In peacetime, perhaps you build wealth (“gold”), while in war, soldiers wear bronze armor. Many sailors died at sea from wrecks, storms, or otherwise, and the sea is vast and uncertain, while the earth represents life. Swords are paired with shields much in the same way that chariots are paired with horses.

Word embeddings are not as broadly interpretable in the way that topic models are, but they can give a sense of the ways in which words are used.

Sentiment Analysis

A lexicon-based approach (using the NRC Emotion Lexicon) was used for sentiment analysis on this corpus. For each word in the lexicon, binary values are assigned for eight emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) as well as for polarity (positive/negative valence).

It’s possible the connotations of some words in these Ancient Greek and Roman works may not necessarily be reflected in the NRC Lexicon, which is one issue I can see with this approach, but it is mostly an exploratory method. The bar chart below provides a summarization of sentiment in each work.



Many works rate highly on fear and anger and have negative polarity. Some exceptions are the *Homeric Hymns*, which rate highly on “joy”; these poems are mostly celebratory of the gods, so this makes sense to some degree.

A more granular way to look at sentiment is to examine a table of sentiments by chapter in a particular work. The table below shows sentiment values for each of the 24 books in the *Iliad*, by Homer.

These sentiments seem only alright at describing emotion in each section. Book 5 (chapter_id 4) rates highly on a lot of emotions, and it depicts the first major battles between the Greeks and the Trojans, so that seems okay. Patroclus, a great friend to Achilles, is killed by Hector in Book 16 (chapter_id 15), and this is also an emotional section according to sentiment analysis.

Some other major events are when Achilles learns of Patroclus' death in Book 18 (chapter_id 17) and when Achilles kills Hector and drags around his corpse in his chariot in Book 22 (chapter_id 21). Neither of these rate particularly highly on emotions like anger or sadness, which doesn't seem entirely correct to me.

		anger	anticipation	disgust	fear	joy	sadness	surprise	trust	polarity
work_id	chapter_id									
9	0	1.730861	1.214255	0.972059	1.817043	1.318384	1.101758	0.438503	1.800535	-0.614109
	1	1.569692	1.009844	0.901147	1.873647	1.077041	1.634125	0.565258	1.986575	-0.609755
	2	1.580227	1.289846	0.814903	1.495365	1.392562	0.990126	0.461849	1.489961	-0.287252
	3	2.181589	1.192394	0.764923	1.963697	1.214398	0.923044	0.488568	1.694840	-0.998156
	4	3.262166	1.549703	1.032995	3.443559	1.272826	1.872256	0.624747	1.753074	-2.597854
	5	1.518744	1.276081	0.829029	1.754722	1.390495	1.209623	0.467575	1.465352	-0.494542
	6	2.123177	1.280881	0.460440	2.238135	1.101224	1.024433	0.431435	1.667092	-1.048688
	7	1.844810	1.103386	0.707958	1.946045	1.180164	1.467844	0.724006	1.595467	-1.019361
	8	1.473605	1.335805	0.995624	1.404845	1.641407	1.443460	0.643026	1.792408	0.043342
	9	1.312702	1.366772	0.599242	1.825061	1.162234	1.134426	0.394226	1.976554	0.082191
	10	2.748826	1.487001	1.016689	2.610600	1.138634	1.887580	0.698926	1.615874	-1.832680
	11	2.160222	0.813262	0.668921	2.208396	0.705244	1.118996	0.405116	1.301246	-1.521955
	12	2.833502	1.647424	0.997055	3.049750	1.244460	1.836388	0.681599	1.801621	-1.829589
	13	1.845478	1.017470	0.496210	1.725659	1.223308	1.368212	0.584145	1.446289	-0.423964
	14	2.487287	1.050446	0.739470	2.572164	0.929834	1.559014	0.530354	1.647100	-1.973406
	15	3.144302	1.535136	1.188081	3.273769	1.128839	2.296683	0.461497	1.583914	-2.597980
	16	2.751813	1.360859	1.373518	2.889983	1.294970	2.315920	0.630066	1.962683	-2.585254
	17	1.522092	1.142675	0.802998	1.861608	1.271713	1.450946	0.799835	1.349188	-0.831561
	18	1.487216	0.979899	0.651299	1.550720	1.145367	1.627072	0.464176	1.503287	-0.637740
	19	2.362125	1.280978	0.688755	2.246929	1.010298	1.151469	0.363628	1.245117	-1.809305
	20	2.050637	1.400315	1.145400	2.378893	1.142389	1.372184	0.508895	1.171902	-1.699190
	21	1.580056	1.369503	0.763420	1.987672	1.169273	1.519195	0.451642	1.288005	-0.991747
	22	1.785274	1.165313	0.943263	1.731714	1.353458	1.600835	0.746350	1.975793	-0.242871
	23	1.976006	1.245570	0.954946	2.422350	1.373259	1.900938	0.653599	1.771831	-0.640793