

SASQUATCH DISTRIBUTION MODELLING

Prof. Joss Wright^{1,2}

¹Oxford Internet Institute
<https://www.oii.ox.ac.uk>

²Weird Data Science
<https://www.weirddatascience.net>

DARK BEGINNINGS

- Tales are whispered throughout human history of unknown human-like creatures in the forests, hills, and wildernesses.
- Despite ravaging human expansion, vast regions of the planet remain darkly secluded from the eyes of civilisation.
- Accumulated accounts, sightings, reports over history accumulate, exposing us to a wealth of data regarding the cryptozoological entities lurking on the edges of our mundane existence.



Gustave Dore - The Enigma. (1871)

TWISTED REFLECTIONS



Unknown - Yeti Painting

- Global oral and written traditions tremulously support the existence of anthropoid relatives of humanity surviving in the remote places of the world.
- Himalayan Yeti
 - ▶ Also known as the Mi-go ("wild man").
 - ▶ Tibetan folklore identifies three types ranging from the 4.5m high black-furred Nyalmo, through the 2.5m Chuti, to the 1.5m high reddish-brown Rang Shim Bombo.
- Yeren of China
 - ▶ Reported in Chinese histories from the Warring States period (475BCE-221BCE) onwards.
 - ▶ Major investigations by the Chinese Academy of Sciences from 1976, following government suppression in the 1950s and 1960s.
- Almaty of Central Asia, Chuchuna of Siberia

TWISTED REFLECTIONS

- Orang Pendek of Indonesia
 - ▶ "Little man". Standing up to 1.5m in height, and with dark grey, black, yellow, or tan fur.
- Yowie of Australia
 - ▶ In indigenous folklore, "a spirit that roams over the earth at night".
 - ▶ Accounts of "Indigenous Apes" appeared in the Australian Town and Country Journal since 1850.
- North American Bigfoot, or Sasquatch.
 - ▶ Attested in Ojibwe petroglyphs - as the Sa'be - dating back several thousand years in the Great Lakes area of North America.
 - ▶ Tsiakto ("wild person") in folklore, otherwise known as "the Night People".



Mathews, A. H. - "A Yaroma
rushed out and swallowed the
Man". (1907)

FORTEAN FOLLOWINGS



Sanderson -
Abominable Snowmen.
p316. (1961)

- Cryptozoology, as a term, was coined by Ivan T. Sanderson.
- Educated at Eton, then Cambridge, Sanderson worked for British Naval Intelligence in World War II, in charge of counter-espionage in the Caribbean.
- Carried out various zoological expeditions in the 1920s and 1930s, and claimed to have been attacked by a cryptid bat - the Olitiau - on an expedition in Cameroon in 1932.
- Led to a lifelong interest in cryptozoological phenomena, including founding the Society for the Investigation of the Unexplained (SITU) in 1967.
- In 1961, published "Abominable Snowmen: Legend Come to Life".
 - Posited four living types of abominable snowmen scattered over five continents.

A FITTING APPEARANCE

- Bigfoot was brought most notably to public attention in 1967 when Roger Patterson and Robert Gimlin captured footage.
- Film was taken in Bluff Creek, California, on the 20th October 1967.
- 954 frames of footage taken on a Cine-Kodak K-100 film camera.
- Shows a dark-haired, humanoid figure, of approximately 2m in height, moving through woodland, turning to look back at the camera.

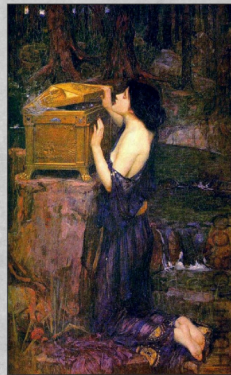


Dore - Cendrillon and the
slipper. From Les Contes de
Ferrault. (1762)



WELLS OF KNOWLEDGE

- Cryptozoological investigations, as with much paranormal research, tends towards a qualitative focus.
 - ▶ Deep interrogation of individual cases, or manual scrutiny of patterns.
- With reckless abandon, we can instead interrogate the phenomenon with the darkest tools of data science, statistics, machine learning.
 - ▶ What are the dread statistical properties of Bigfoot sightings?
 - ▶ What horrific truths can we encounter, if we shine the cold precision of mathematics into the arboreal shadows of the world?



John William Waterhouse -
Pandora (1896)

DREDGING THE DEPTHS



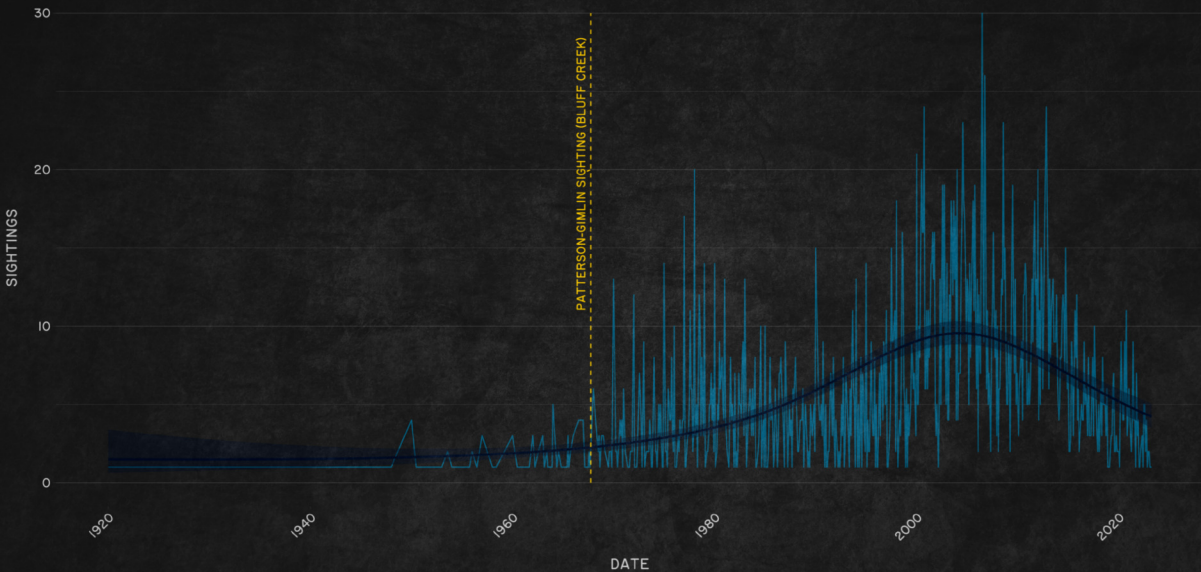
Gustave Doré - The Deluge

- The Bigfoot Field Researchers Organization^a maintain an online database of reported sightings in the US and Canada.
- Sightings are reported manually, and recorded with date, specified location, and details of the sighting.
- For our arcane purposes, a system of retrieval is necessary.
 - ▶ Each individual sighting retrieved from the website, and collated into a database.
 - ▶ Locations parsed and geolocated to give geographical coordinates.
 - ▶ Once stripped of corrupted reports, results in a database of 3,892 sightings spanning from 1920 onwards.

^a <https://www.bfro.net>

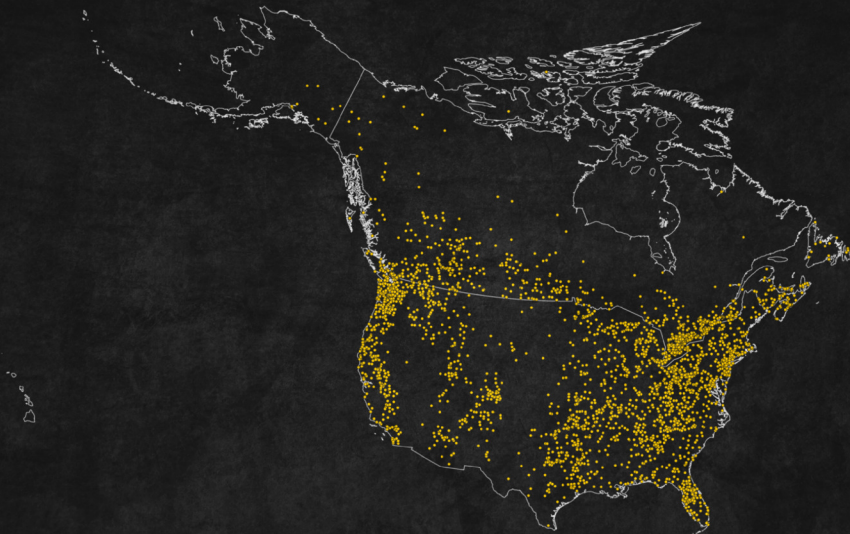
SASQUATCH SIGHTING TIMELINE

<https://www.weirddatascience.net>



SASQUATCH SIGHTINGS // NORTH AMERICA

<http://www.weirddatascience.net> | [WeirdDataSci](#)



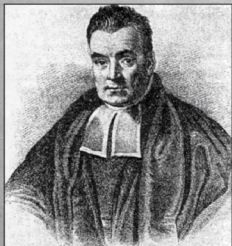
INFERENCE STATISTICS

- Simple descriptive statistics can only take us so far in our descent into madness.
- Inferential statistics twist and distort raw observations to predict future phenomena.
- Two major schools of inferential statistics.
 - ▶ Frequentist statistics have dominated the last century,
 - ▶ Bayesian statistics provide an increasingly popular alternative.



Eugene Delacroix - Lycurgus Consulting the Pythia. (c.1840)

BAYES THEOREM



Rev. Thomas Bayes

Bayes' theorem is a tool to update our beliefs in the presence of data.

$$P(\theta|\text{Data}) = \frac{P(\text{Data}|\theta)P(\theta)}{P(\text{Data})}$$

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Fundamental result in probability theory; core to Bayesian inference.

BAYESIAN WHISPERINGS

- Bayesian inference treats unknowns as probability distributions, not fixed, unknown quantities.
 - ▶ Easy incorporation of new data - a posterior becomes a prior as new evidence arrives.
 - ▶ Results emerge as full probability distributions over chilling unknowns, not point estimates.
 - ▶ Lends itself to hierarchical models, in which correlated phenomena become locked in insidious mutual influence.



Albrecht Dürer - Melancholia

TOOLS

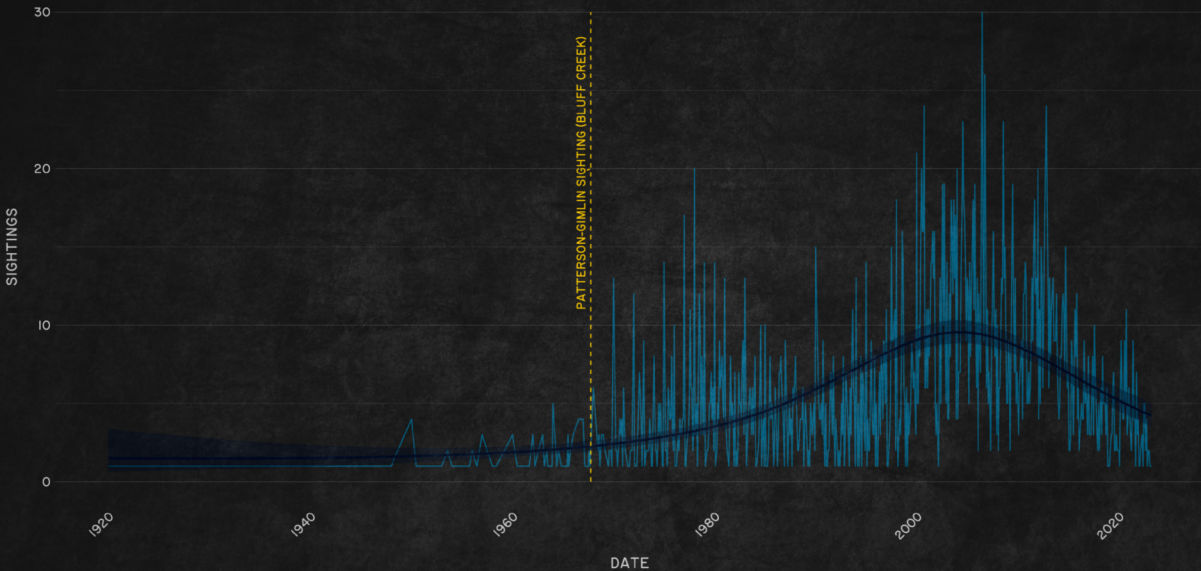


Daret. after Blanchard. Vulcan
forging the armour of Achilles.
(1624-1675)

- Significant computational effort is required to determine the evidence, or "marginal likelihood", term.
- Computational approaches fit models by arcane explorations of the multidimensional space of hyperparameters.
- Probabilistic programming languages convert a statistical model to code.
 - Stan: <https://www.mc-stan.org>

SASQUATCH SIGHTING TIMELINE

<https://www.weirddatascience.net>



TEMPORA MUTANTUR



Gustave Doré - The Adventure
with the Windmills. from
Illustrations to Don Quixote.
(1863)

- To interrogate the behaviours of cryptid species, and inform our later analyses, carry out a Bayesian changepoint analysis.
- Identify notable shifts in Bigfoot behaviour - are there points in history where sightings have notably increased?
- Why might this be?

DEFINED CONSTRAINTS

- Consider sightings as drawn from a `_negative binomial_` distribution.
- The number of sightings in a given month is the number of successful trials.
- Model the series of sightings as `_two_` negative binomial distributions, with discrete parameters.
 - ▶ At what point in time do two fitted negative binomial distributions split such that they best describe the series?
- Reveals the most significant point of statistical change in the historical data.



Dore - The Inferno. Canto XXXIV. Satan.
trapped in the frozen central zone in the
Ninth Circle of Hell.

TEMPORAL FRACTURES



Giovanni Battista Piranesi -
Carceri. Folder 7 (1745)

Formally:

$$\Omega \sim \text{NegBinomial}((t < c? \mu_e, \mu_1), (t < c? \phi_e^{-1}, \phi_1^{-1}))$$

$$\mu_e, \mu_1 \sim \text{HalfCauchy}(5)$$

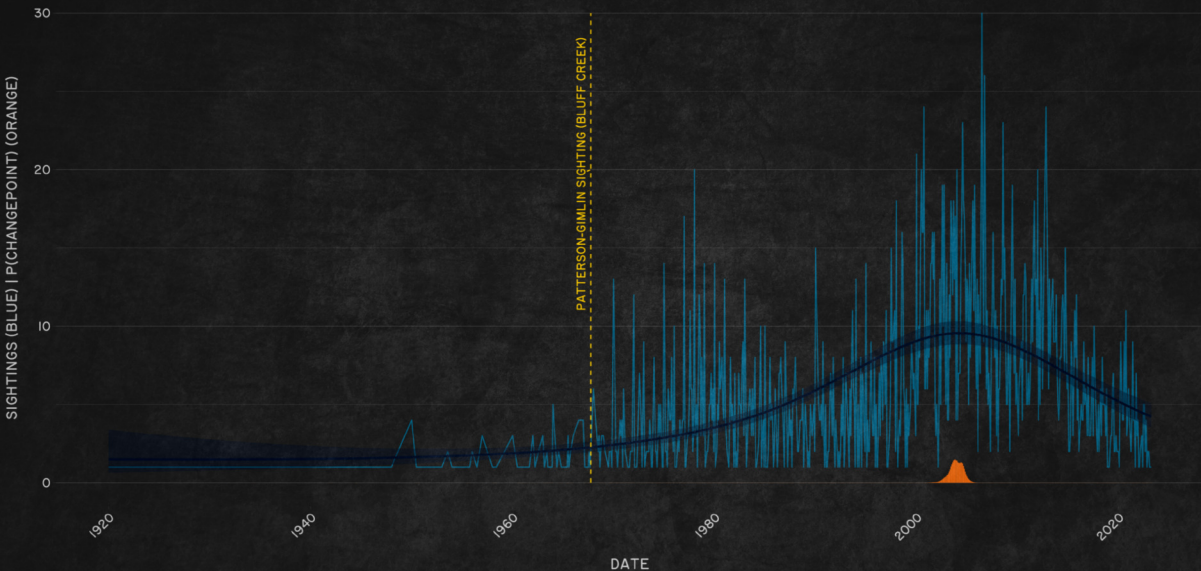
$$\phi_e^{-1}, \phi_1^{-1} \sim \text{HalfCauchy}(1)$$

$$c \sim \text{Uniform}(1, t)$$

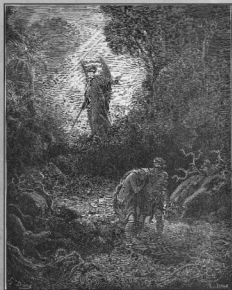
Implemented in the Stan probabilistic programming language for Bayesian model fitting and assessment.

SASQUATCH SIGHTING TIMELINE // CHANGEPOINT PROBABILITY

<https://www.weirddatascience.net>



SHATTERED COHERENCE



Dore - Adam and Eve Driven Out
of Eden (1866)

- The model identifies a relatively coherent shift in sightings around 2003-2004.
- Modal probability of the significant changepoint occurs in November 2003.
- For the moment, we will store this unsettling insight to be placed later into its full, squalid context, and turn what remains of our minds to spatial phenomena.

CAELUS NOCTURNUS

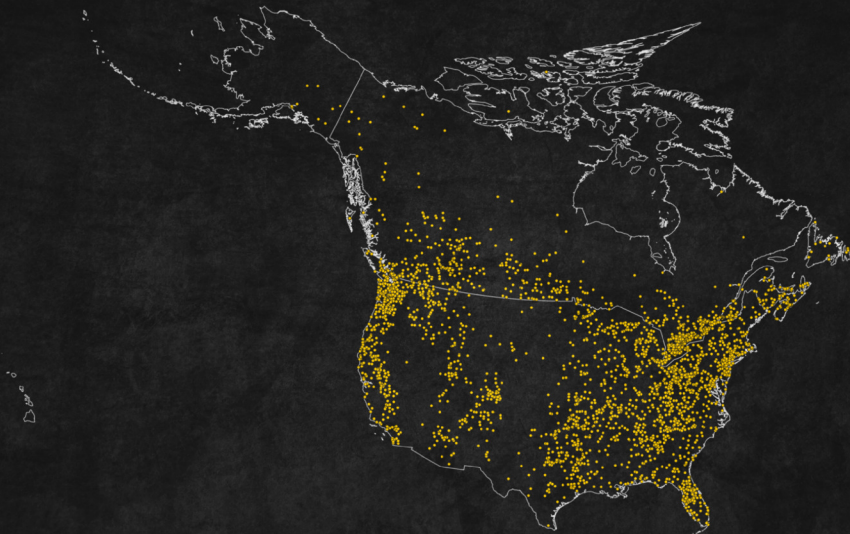
- Having considered time, we turn our attention to space.
- What can we learn from the shifting patterns of sightings spread across the forests, plains, and mountains of North America?
- Where do sasquatch congregate? What draws them, ever elusive, to particular regions?
- We turn, irrevocably, to species distribution modelling.



atlas bearing the heavens in the
form of an armillary sphere.
from William Cunningham. The
Cosmographicaall Glasse. London
1559.

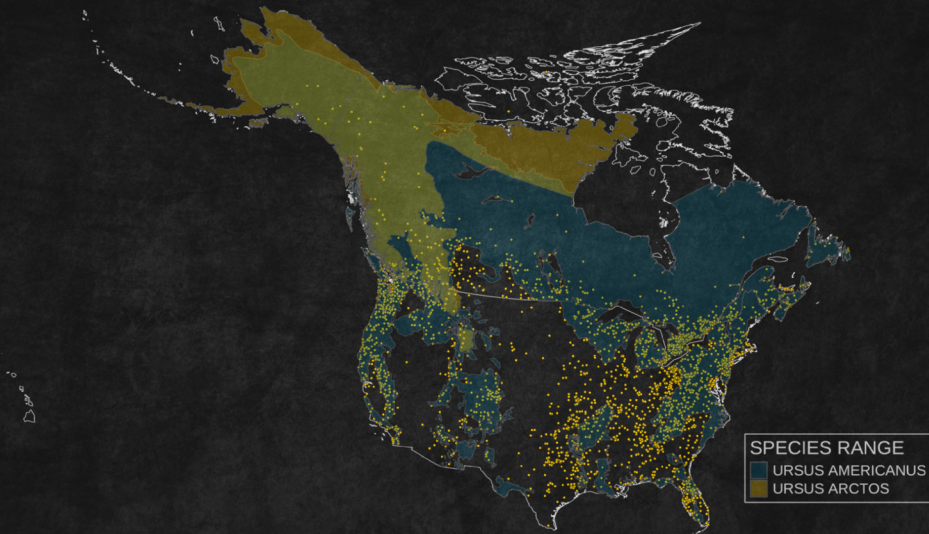
SASQUATCH SIGHTINGS // NORTH AMERICA

<http://www.weirddatascience.net> | [WeirdDataSci](#)

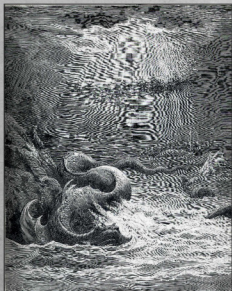


SASQUATCH SIGHTINGS // URSUS AMERICANUS, URSUS ARCTOS RANGES

<http://www.weirddatascience.net> | WeirdDataSci



SEMANTIC MUTATIONS

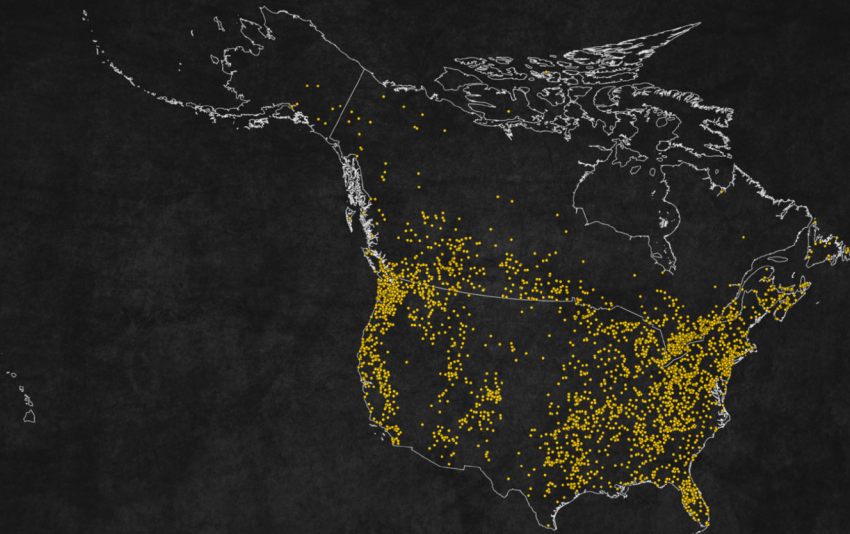


Dore - The Creation of Fish and
Birds (1868)

- To understand the key habitats for Bigfoot, we must enrich our pitifully sparse data.
 - ▶ Understand what hidden factors unite, and eternally divide, sightings.
 - ▶ Temperature, elevation, forest cover, human population density, crop cover, grass cover, precipitation.
 - ▶ Using geolocated sightings data, mark each point with `_bioclimatic variables_` selected according to our intuitions.
- The tragically, and suspiciously, closed-minded nature of research funding restricts us to opportunistic sightings.
 - ▶ Without confidence in missing data, we create an even distribution of 'pseudo-absence' points spanning the geographical range.
 - ▶ Adding bioclimatic data to these provides needed contrast.

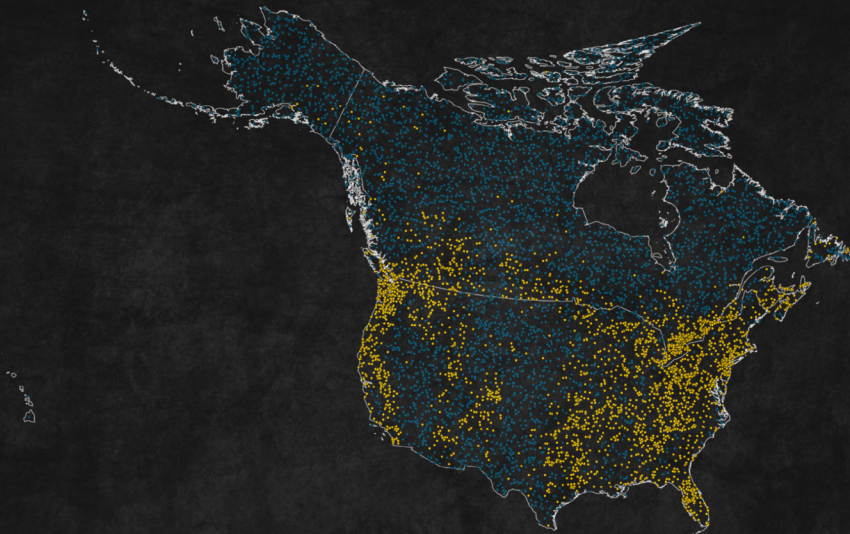
SASQUATCH SIGHTINGS // NORTH AMERICA

<http://www.weirddatascience.net> | [WeirdDataSci](#)



SASQUATCH SIGHTINGS // PSEUDO-ABSENCES

<http://www.weirddatascience.net> | [WeirdDataSci](#)



HIDDEN LOGISTICS

- Model presence or absence as the outcome of a generalized linear model (GLM).
 - ▶ Specifically, a `_logistic regression_`, predicated on the bioclimatic variables.
- Models the log-odds of a sighting occurring at a given location, given its climatic features.



Dore - The Inferno. Canto XXXIV. Satan.
trapped in the frozen central zone in the
Ninth Circle of Hell.

SIGMOIDAL UNDULATIONS



Gustave Moreau - Heracles and
the Lernaean Hydra

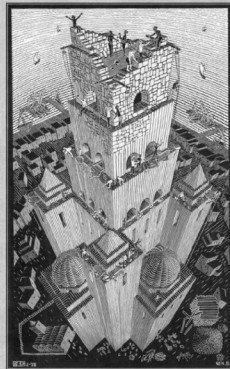
Formally:

$$\begin{aligned}y &\sim \text{Bernoulli}(\phi) \\ \text{logit}(\phi) &= \alpha + \beta x \\ \alpha &\sim \mathcal{N}(0, 1) \\ \beta &\sim \mathcal{N}(0, 1)\end{aligned}$$

Across our range of bioclimatic variables.

TOPOLOGICAL MUTABILITES

- When dragging spatial features into our previously abstract and directionless world, we face the problem of _spatial autocorrelation_.
- Models typically assume that each point is unrelated to each other point.
- Tobler's First Law of Geography:
 - ▶ "Everything is related to everything else, but near things are more related than distant things."
- Geographically close points are more likely to be correlated, and should thus contribute less information to the model.



Escher - Tower of Babel

NEPHILIM FIELDS

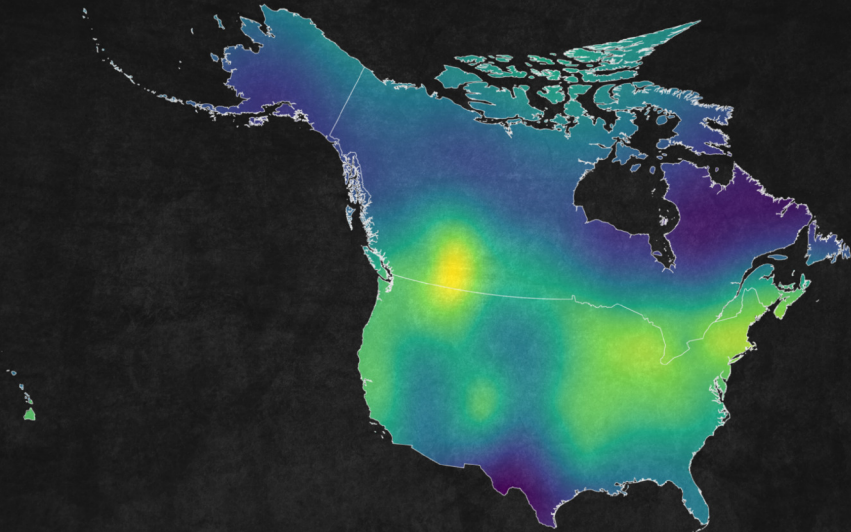


Dore - Paradise Lost (Wilton).
Michael casts out rebel angels.
(1866)

- Accounting for spatial interactions massively increases computational requirements, making it infeasible in general for standard Bayesian tools at large scale.
- An approximation mechanism, the `_integrated nested laplace approximation_` (INLA) reduces computational requirements at a small cost of accuracy, under certain assumptions.
- Extract the underlying `_spatial field_` from sightings, revealing the horrible truth of the covariates.
- Estimated using a stochastic partial differential equation (SPDE) via log-Gaussian Cox processes (LGCP).

SASQUATCH MODEL SPATIAL FIELD

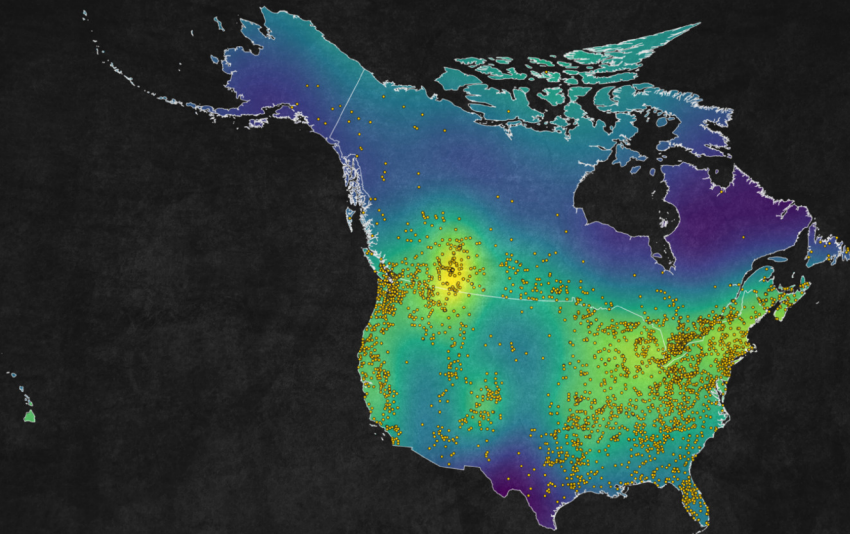
<http://www.weirddatascience.net> | WeirdDataSci



Data: <https://www.bfro.net>

SASQUATCH MODEL SPATIAL FIELD // SIGHTINGS OVERLAP

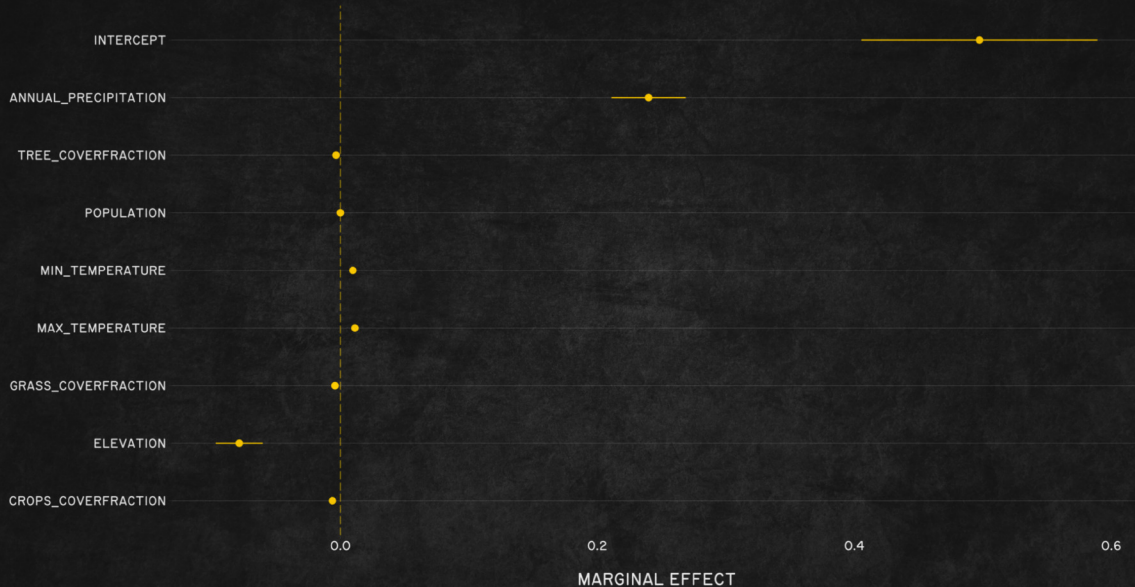
<http://www.weirddatascience.net> | WeirdDataSci



SASQUATCH MODEL COVARIATES

<http://www.weirddatascience.net> | [WeirdDataSci](#)

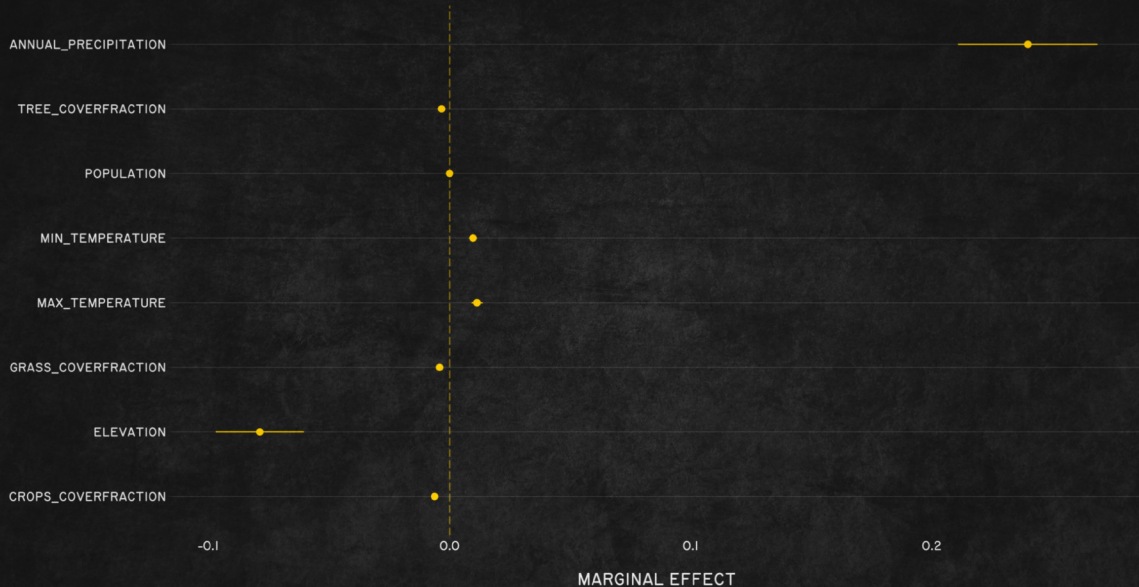
PREDICTOR



SASQUATCH MODEL COVARIATES // NO INTERCEPT

<http://www.weirddatascience.net> | [WeirdDataSci](#)

PREDICTOR



TERRA COGNITA

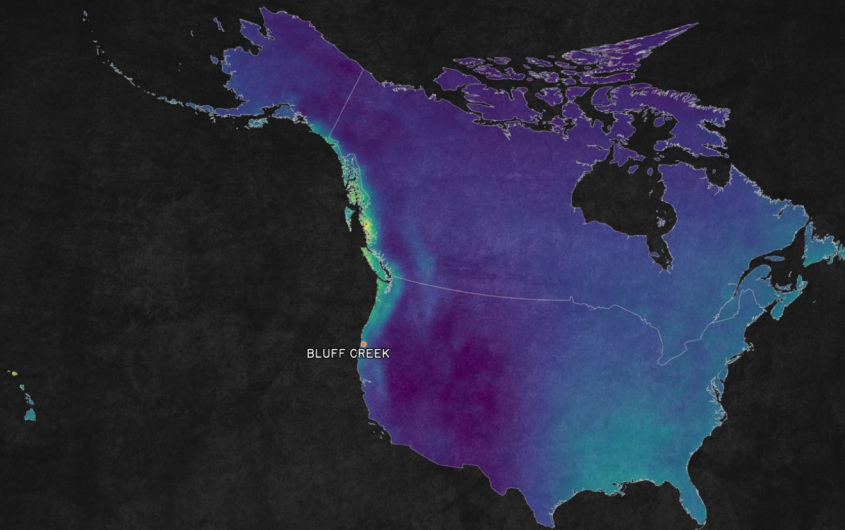
- The environmental variables, absent the underlying spatial effects, allow a predictive assessment of habitat suitability.
 - ▶ What swirling combination of conditions best suit the sasquatch?
- Willfully extract bioclimatic variables across North America to recast the model inexorably to its underlying geography.



Dore - Dante's Inferno. Virgil
pushes Filippo Argenti back into
the river Styx. (1890)

SASQUATCH HABITAT SUITABILITY // NORTH AMERICA

<https://www.weirddatascience.net>



Data: <https://www.bfro.net>

PRECIPITOUS ASSUMPTIONS

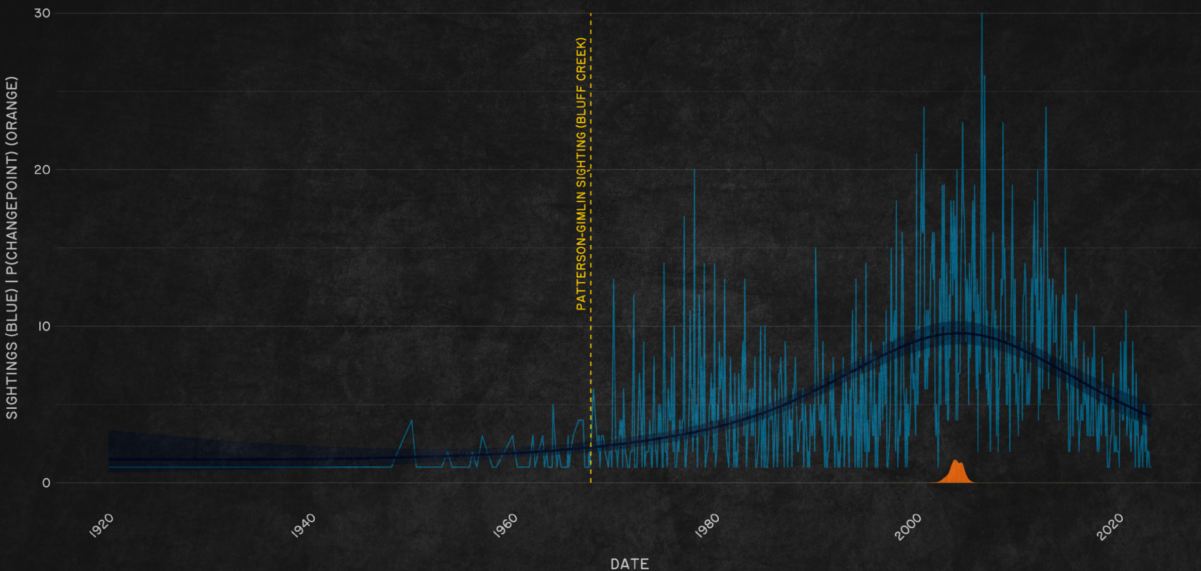
- Drifting further from mundane reality - does this model shatter our preconceptions concerning the timeline of sightings?
- With precipitation revealed as a significant factor in sasquatch suitability, what conclusions might we wrest from the void?



Dore - The Inferno. Canto XXI.
"Be none of you outrageous."

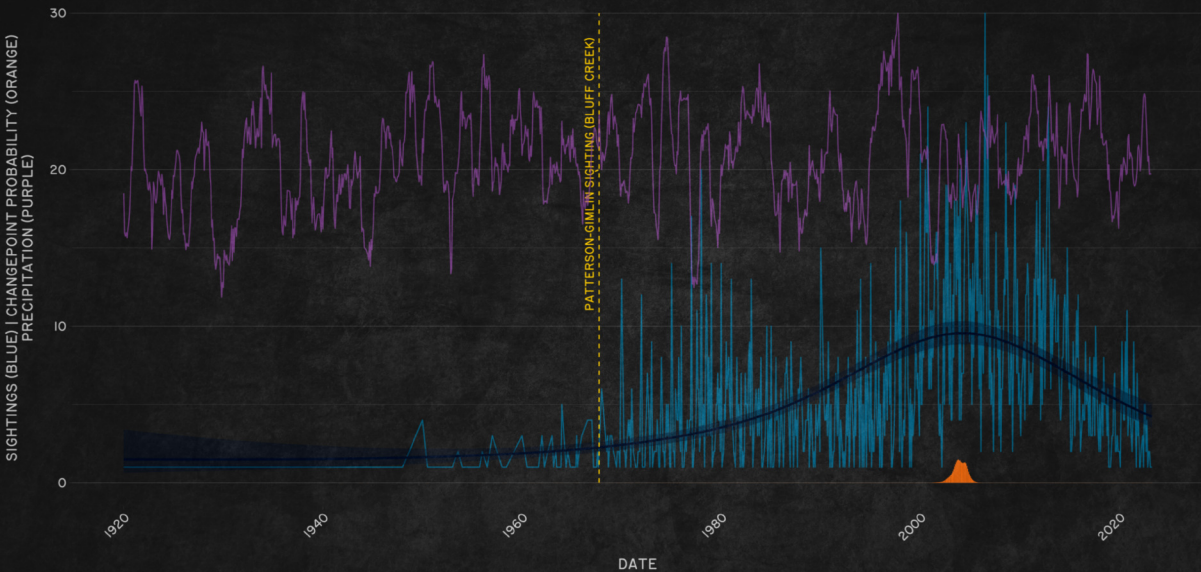
SASQUATCH SIGHTING TIMELINE // CHANGEPOINT PROBABILITY

<https://www.weirddatascience.net>



SASQUATCH SIGHTING TIMELINE // CHANGEPOINT // PRECIPITATION

<https://www.weirddatascience.net>



DELPHIC REVELATIONS

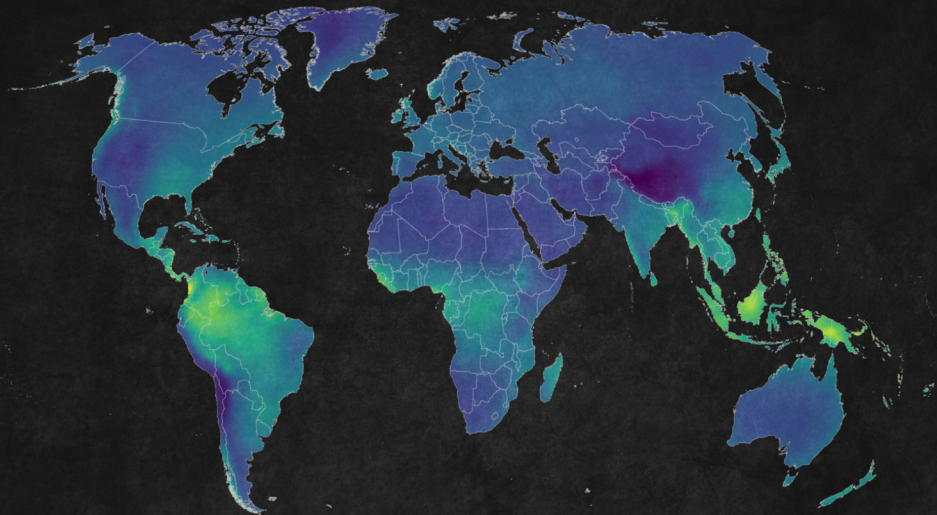


Jan Swart van Croningen.
Cassandra predicts the downfall
of Troy. (1550-1555)

- Unbound by constraints, we now have a model that can link bioclimatic variables to species habitat suitability for the North American Sasquatch.
- Extend our suitability to the global scale.
 - ▶ Sample a regular grid of points across the entire globe, and retrieve appropriate bioclimatic variables.

SASQUATCH HABITAT SUITABILITY // GLOBAL

<https://www.weirddatascience.net>



TERRA OCCULTA

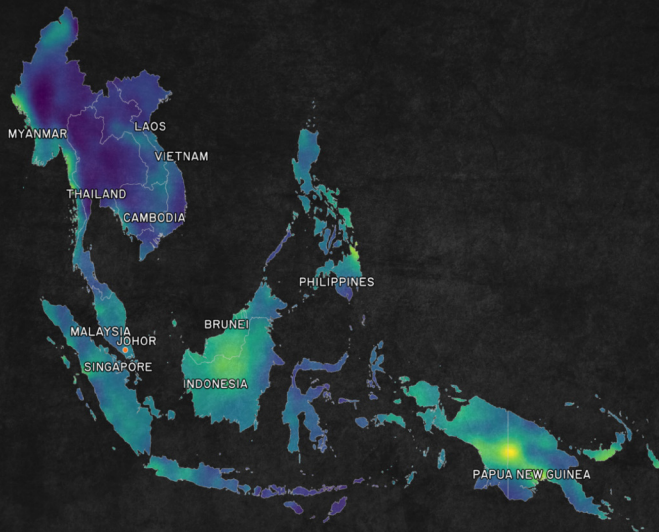


Friedrich Preller the
Elder. Calypso Abschied
von Odysseus. (1864)

- A chill demarcation, supporting Sanderson's wildest conjectures, of different sub-species of Sasquatch
 - ▶ The Yeti ranges of the Himalayas and Tibetan Plateau is utterly unsuited to the North American Sasquatch.
 - ▶ The rumoured territories of Central Asian Almasty, the Australian Yowie, and the Chinese Yeren are more congenial, but ultimately lacking.
- Where else do we see nascent stirrings of potential climatic conjuncture?
 - ▶ Exemplary circumstances appear to fall within South America, most notably Colombia.
 - ▶ Southeast Asia presents a promising vista of potential habitability.

SASQUATCH HABITAT SUITABILITY // SOUTHEAST ASIA

<https://www.weirddatascience.net>



In search of Bigfoot, scientists may uncover unknown biodiversity in Malaysia



RHETT A. BUTLER

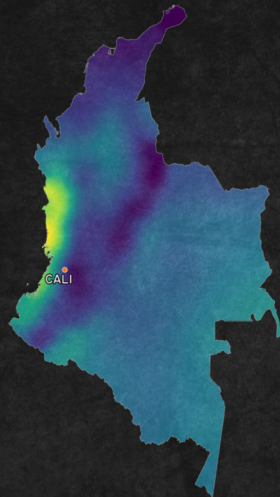
1 FEB 2006 MALAYSIA

Malaysian scientists are scouring the rainforests of Johor state in search of the legendary ape-man Bigfoot, supposedly sighted late last year. But they are more likely to encounter some less fantastic but unique creatures that dwell in these still unexplored ecosystems.

In recent years a number of new and conspicuous animals have been discovered in the forests of South and Southeast Asia, including [361 new species](#) in Borneo over the past decade and 43 new species of vertebrates in Sri Lanka. Just last year, scientists with the World Wildlife Fund (WWF) captured on film what may be a new species of mammal in the rainforests of Kalimantan in Indonesian Borneo. The [fox-like creature](#) is apparently unknown even to local hunters.

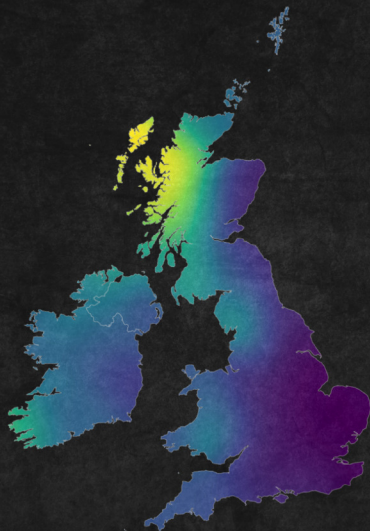
SASQUATCH HABITAT SUITABILITY // COLOMBIA

<https://www.weirddatascience.net>



SASQUATCH HABITAT SUITABILITY // BRITISH ISLES

<https://www.weirddatascience.net>



Data: <https://www.bfro.net>

DISTURBING CONCLUSIONS

- A century of Sasquatch sightings reveals its most likely habitat to be the traditional Pacific Northwest.
- The North American Sasquatch favours warm, low-lying, areas, with high precipitation.
 - ▶ High rainfall potentially associated with greater activity.
- Global folklore concerning related cryptids suggests various species co-existing globally in the shadows of humanity's encroaching disruptions of the natural world.
- Despite everything we have learned, perhaps the real terror is the tortured mathematics we performed on the journey.



Francisco Jose de Goya y
Lucientes - "The Sleep of Reason
Produces Monsters" from "Los
Caprichos". (1797-1799)