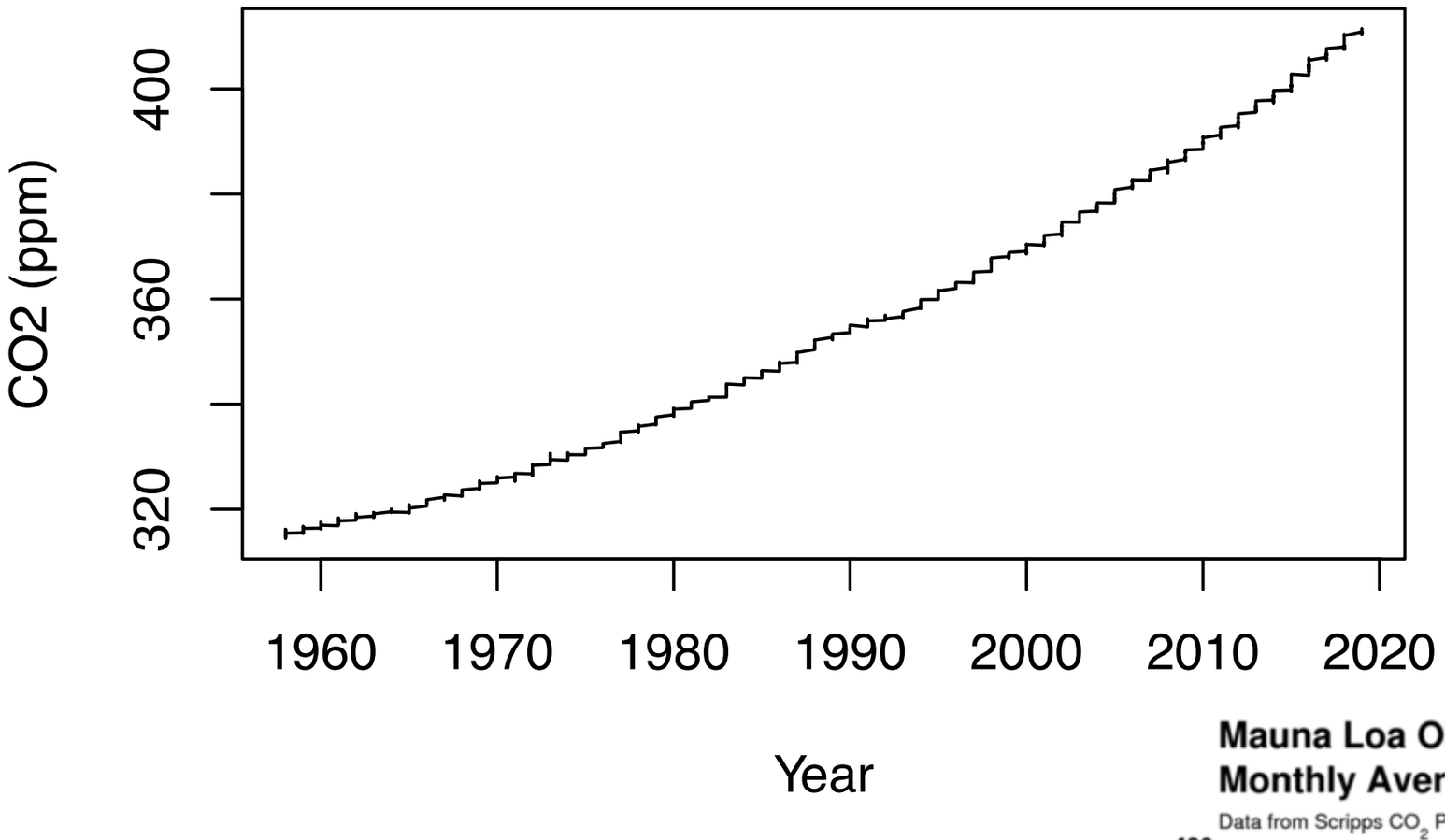
Lecture 2.1 – Finding Meaning in Data

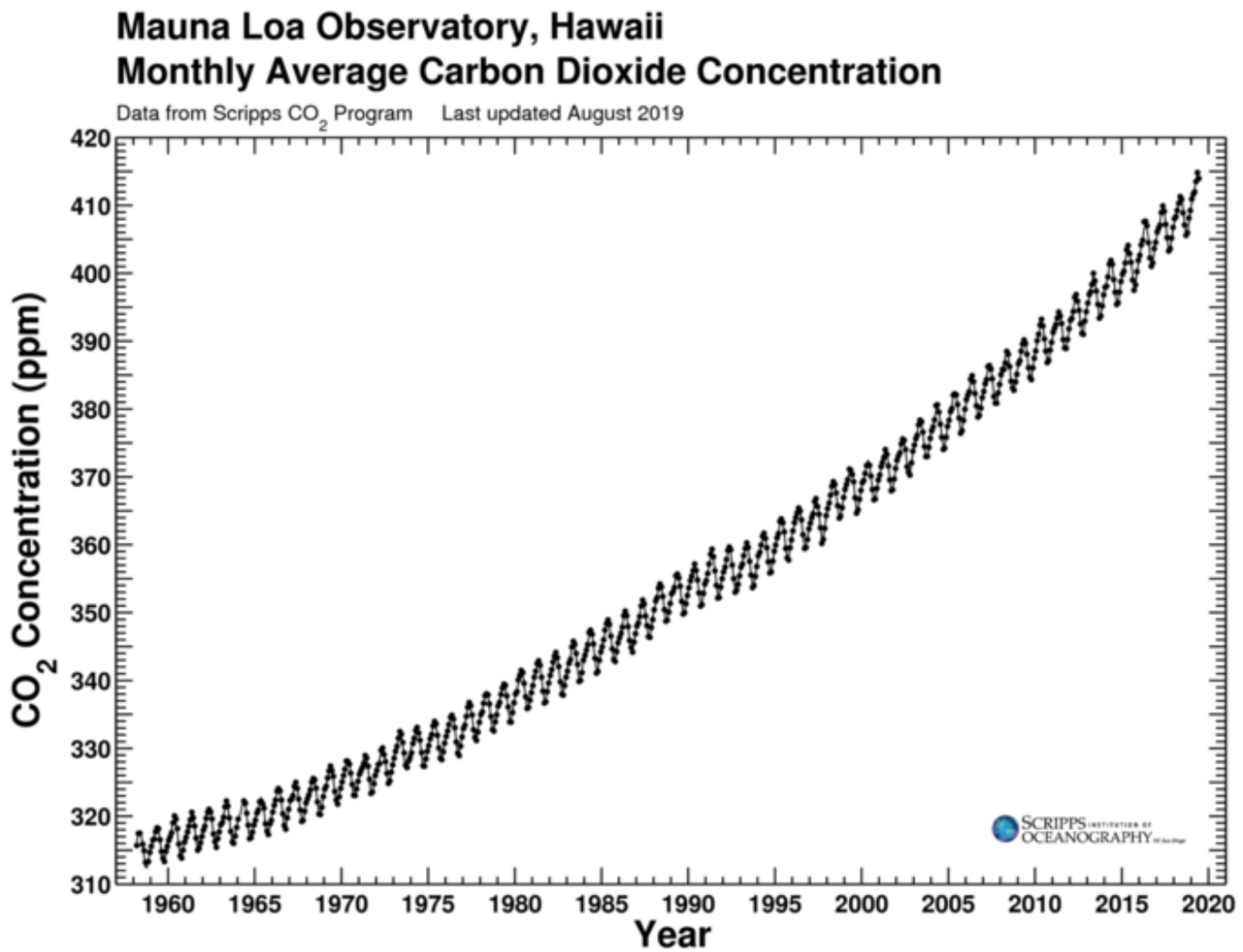
Specific Learning Objectives:

3. Independently perform basic data analysis and visualizations in a way that communicates ideas clearly.

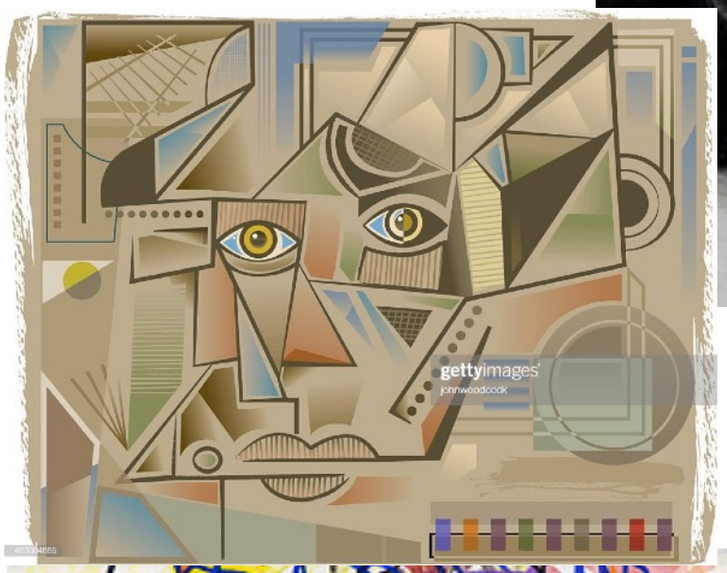
2003	4	37726	2003.2877	377.65	374.98	377.82	375.13	377.65	374.98
2003	5	37756	2003.3699	378.35	375.11	378.58	375.33	378.35	375.11
2003	6	37787	2003.4548	378.13	375.67	377.97	375.54	378.13	375.67
2003	7	37817	2003.5370	376.60	375.82	376.48	375.73	376.60	375.82
2003	8	37848	2003.6219	374.48	375.94	374.43	375.92	374.48	375.94
2003	9	37879	2003.7068	372.98	376.32	372.74	376.10	372.98	376.32
2003	10	37909	2003.7890	373.00	376.46	372.82	376.27	373.00	376.46
2003	11	37940	2003.8740	374.35	376.51	374.29	376.43	374.35	376.51
2003	12	37970	2003.9562	375.69	376.58	375.70	376.57	375.69	376.58
2004	1	38001	2004.0410	376.79	376.75	376.76	376.71	376.79	376.75
2004	2	38032	2004.1257	377.37	376.64	377.57	376.84	377.37	376.64
2004	3	38061	2004.2049	378.39	376.89	378.48	376.96	378.39	376.89
2004	4	38092	2004.2896	380.50	377.80	379.80	377.08	380.50	377.80
2004	5	38122	2004.3716	380.62	377.36	380.44	377.19	380.62	377.36
2004	6	38153	2004.4563	379.55	377.11	379.72	377.31	379.55	377.11
2004	7	38183	2004.5383	377.76	377.00	378.16	377.43	377.76	377.00
2004	8	38214	2004.6230	375.83	377.32	376.04	377.57	375.83	377.32
2004	9	38245	2004.7077	374.05	377.41	374.34	377.72	374.05	377.41
2004	10	38275	2004.7896	374.22	377.69	374.43	377.88	374.22	377.69
2004	11	38306	2004.8743	375.84	378.01	375.92	378.06	375.84	378.01
2004	12	38336	2004.9563	377.44	378.33	377.37	378.25	377.44	378.33
2005	1	38367	2005.0411	378.34	378.30	378.50	378.45	378.34	378.30
2005	2	38398	2005.1260	379.61	378.88	379.40	378.66	379.61	378.88
2005	3	38426	2005.2027	380.08	378.60	380.36	378.86	380.08	378.60
2005	4	38457	2005.2877	382.05	379.36	381.78	379.08	382.05	379.36
2005	5	38487	2005.3699	382.24	378.98	382.55	379.29	382.24	378.98
2005	6	38518	2005.4548	382.08	379.61	381.96	379.51	382.08	379.61
2005	7	38548	2005.5370	380.66	379.87	380.48	379.72	380.66	379.87
2005	8	38579	2005.6219	378.67	380.13	378.44	379.94	378.67	380.13
2005	9	38610	2005.7068	376.42	379.78	376.77	380.15	376.42	379.78
2005	10	38640	2005.7890	376.80	380.28	376.88	380.35	376.80	380.28
2005	11	38671	2005.8740	378.31	380.48	378.39	380.55	378.31	380.48
2005	12	38701	2005.9562	379.96	380.85	379.86	380.73	379.96	380.85
2006	1	38732	2006.0411	381.37	381.32	380.97	380.91	381.37	381.32
2006	2	38763	2006.1260	382.02	381.29	381.82	381.09	382.02	381.29
2006	3	38791	2006.2027	382.56	381.07	382.74	381.23	382.56	381.07
2006	4	38822	2006.2877	384.36	381.67	384.10	381.39	384.36	381.67
2006	5	38852	2006.3699	384.92	381.65	384.81	381.53	384.92	381.65

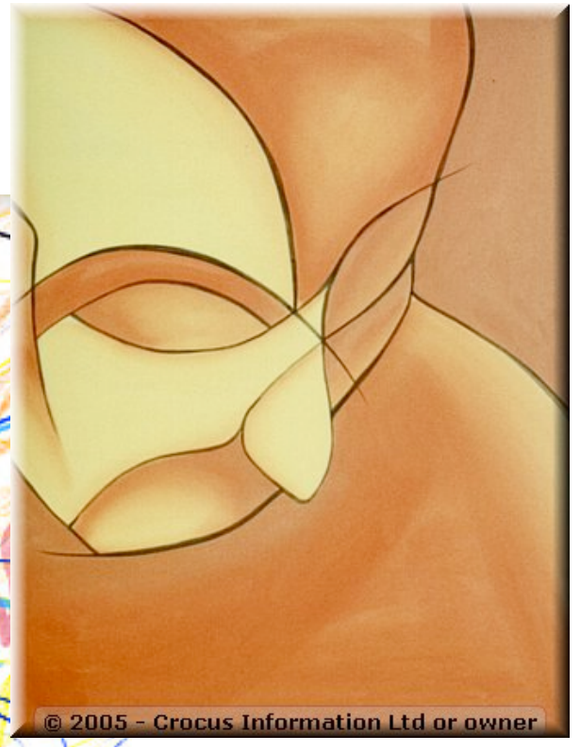


- Finding meaning in data means finding meaningful patterns.
- Humans find patterns quickly in visual data presentation.
- Data visualization is a very important part of data analysis and communication!



Visualization of a concept in art





Visual communication is highly efficient!

There are many ways of communicating ideas clearly!

In this course, we will learn how to explore data

- We've already learned some tools to help deal with large data sets (R), and we'll continue to build on those.

- We'll learn more about the importance of sight and how the brain detects and understands visual patterns.

- You'll have a chance to work with a data set on a topic that you find interesting!

Project 1: Visualization of a Data Set

- This A-level work will require you to handle a data set in R, create a meaningful visualization of results, and present it to the class for critique.

- **First**: choose a data set based on what you are interested in. It doesn't matter what topic, just whatever you want to dive into!

Data set ideas have to be approved by Oct 4 at 5 pm!

- **Second**: Find a pattern. Do some visualizations, explore the data, and come up with ideas about what the data are telling you!

- **Third**: Make a visualization. Try to communicate your idea clearly using a visualization!

Project 1: Visualization of a Data Set

Ideas of where to look:

Organization	Website	Topics
FiveThirtyEight	https://data.fivethirtyeight.com/	sports, politics, polling data, gaming
New York Times	https://github.com/nytimes	COVID, historical events
Kaggle	https://www.kaggle.com/	Machine learning, marketing, pop culture
Figshare	https://figshare.com/	science, mathematics

Data set ideas have to be approved by Oct 7 at 5 pm!

Action Items

1. **Choose a data set to work with for Project 1.**

2. **Read Davies Chapter 7.1-7.3 and Chang Chapters 1-2 for next time.**