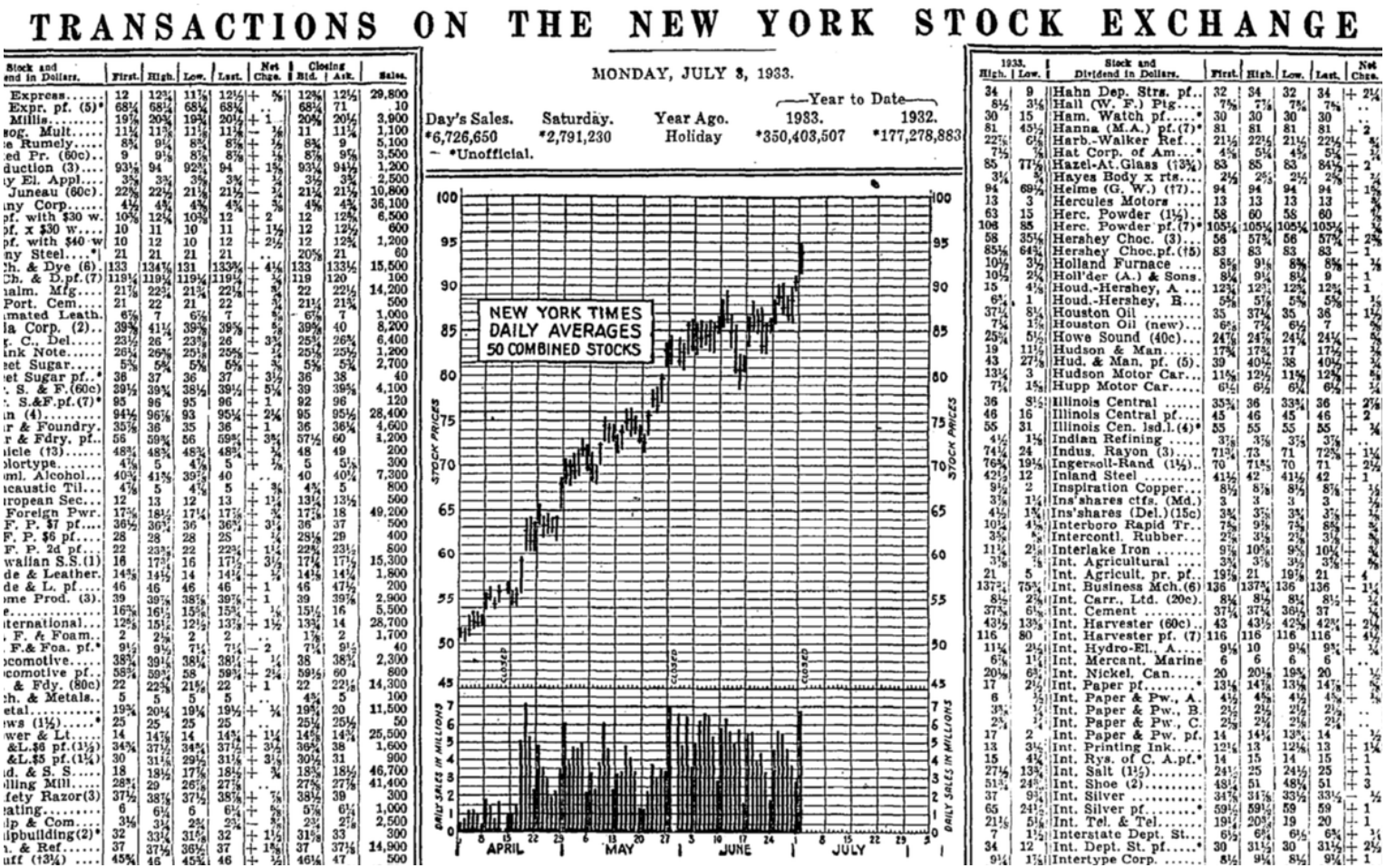
Introduction to data science & artificial intelligence (INF7100)

Arthur Charpentier

#291 Time Series

été 2020

Time Series



TRANSACTIONS ON THE NEW YORK STOCK EXCHANGE

MONDAY, JULY 3, 1933.

Day's Sales.	Saturday.	Year Ago.	Year to Date 1933.	Year to Date 1932.
*6,726,650	*2,791,230	Holiday	*350,403,507	*177,278,883

— *Unofficial.

NEW YORK TIMES DAILY AVERAGES 50 COMBINED STOCKS

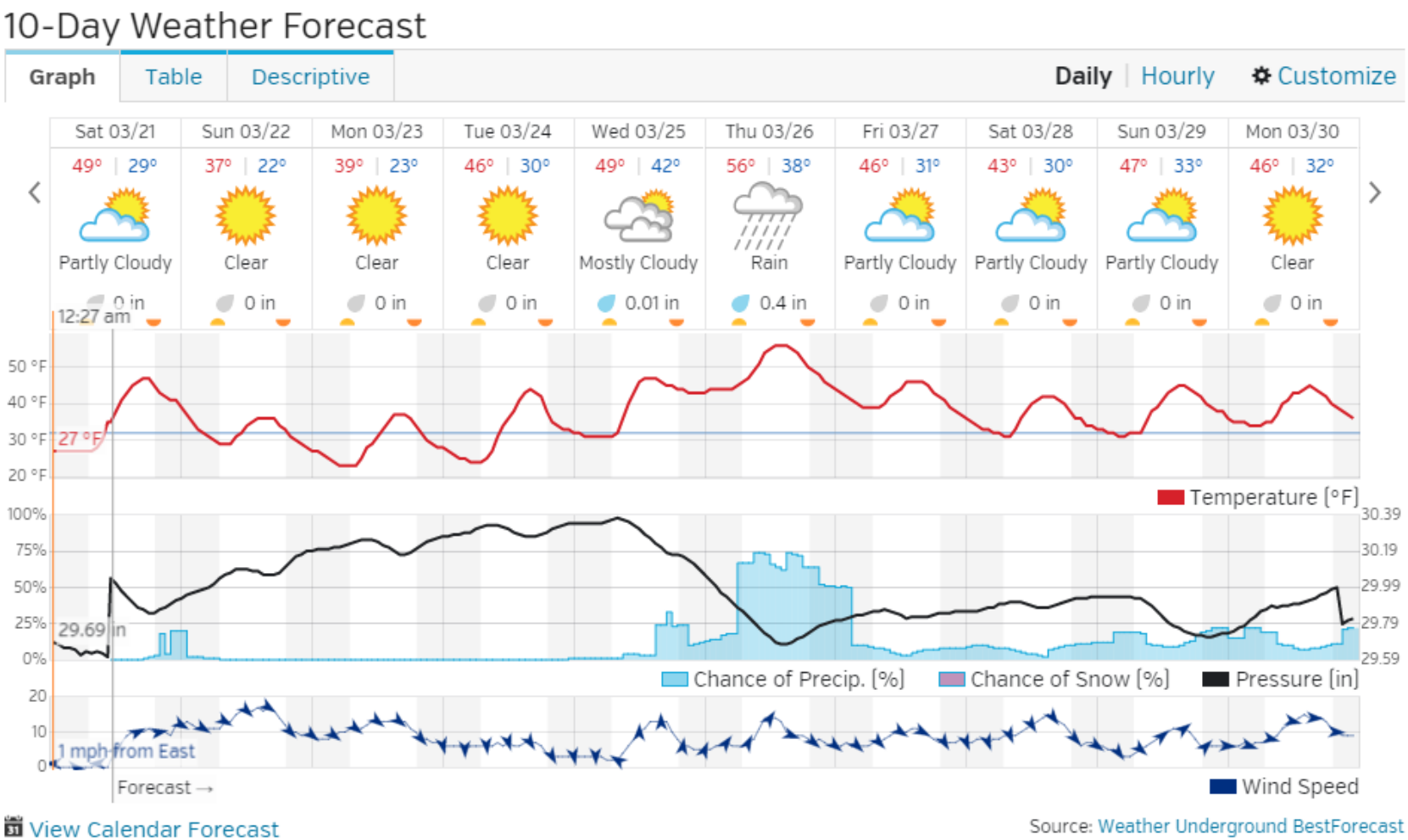
STOCK PRICES

DAILY SALES IN MILLIONS

APRIL MAY JUNE JULY

via When Did Charts Become Popular?

Time Series



via http://stat4701.github.io/

Time Series

A time series is a sequence of observations (Y_t) ordered in time, at regular dates.

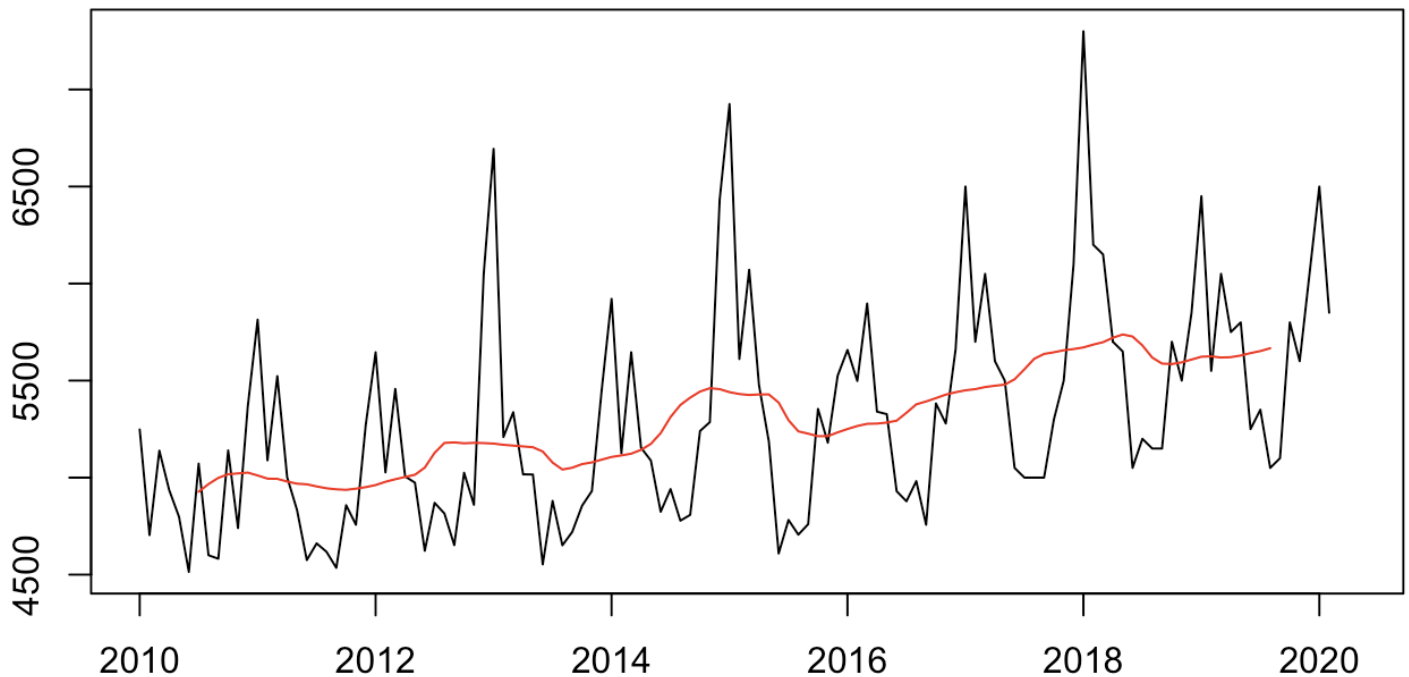
Buys-Ballot (1847, Les changements périodiques de température, dépendants de la nature du soleil et de la lune, mis en rapport avec le pronostic du temps, déduits d'observations néerlandaises de 1729 à 1846) - original probably in Dutch.

TABLEAU REPRÉSENTANT LA MARCHE DE LA TEMPÉRATURE PENDANT L'ANNÉE.

Date.	Temp.	Temp. calculée.	Différ.	Temp. calculée.	Différ.	Date.	Temp.	Temp. calculée.	Différ.	Temp. calculée.	Différ.
10 Janv.	32.58	32.58	0	32.58	0	17 Juill.	0.58	+ 0.24	+ 0.34	63.68	+ 0.66
15 «	+ 1.12	+ 0.78	+ 0.23	+ 0.88	+ 0.13	22 «	— 0.36	64.33	— 0.25	64.03	+ 0.05
20 «	0.69	34.14	+ 0.25	34.34	+ 0.05	27 «	+ 0.50	64.57	+ 0.01	64.37	+ 0.21
25 «	0.83	34.92	+ 0.30	35.22	0	1 Août	+ 0.26	64.83	+ 0.02	64.71	+ 0.13
30 «	1.26	35.70	+ 0.78	+ 0.59	+ 0.67	6 «	+ 0.22	65.06	0	65.06	0
4 Févr.	— 0.01	36.47	0	36.40	+ 0.07						
9 «	+ 0.39	+ 0.55	— 0.16	36.99	— 0.13	11 «	— 0.45	— 0.42	— 0.03	— 0.50	+ 0.05
14 «	+ 0.36	37.57	— 0.35	37.58	— 0.36	16 «	0.34	64.22	+ 0.05	64.05	+ 0.22
19 «	— 0.05	38.13	— 0.96	38.17	— 1.00	21 «	0.42	63.80	+ 0.05	63.55	+ 0.30
24 «	+ 1.59	38.68	+ 0.08	38.76	0	26 «	0.66	63 38	+ 0.19	63.04	+ 0.15
1 Mars	0.48	39.24	0	+ 0.69	— 0.25	1 Sept.	0.23	62.96	0	62.54	+ 0.42
6 «	0.25	+ 0.73	— 0.48	40.14	— 0.65	6 «	0.93	— 1.13	+ 0.20	62.03	0
11 «	0.11	40.69	— 1.10	40.83	— 1.23	11 «	1.12	60.71	+ 0.06	— 1.26	0
17 «	1.83	41.42	+ 0.03	41.52	— 0.09	16 «	0.89	59.59	+ 0.43	59.51	+ 0.51
22 «	0.79	42.15	+ 0.07	42.22	0	21 «	1.37	58.46	+ 0.19	58.25	+ 0.41
27 «	0.67	42.89	0	+ 1.29	— 0.62	26 «	1.32	57.33	0	56.99	+ 0.34
1 Avril	1.34	+ 1.36	— 0.02	44.80	— 0.57	1 Oct.	1.59	— 1.32	0.27	55.74	0
6 «	1.90	45.61	+ 0.52	46.09	+ 0.04	6 «	1.08	54.69	— 0.03	— 1.43	+ 0.35
11 «	1.23	46.99	+ 0.37	47.38	— 0.02	11 «	1.49	53.37	— 0.20	52.88	+ 0.29
16 «	0.89	48.36	— 0.11	48.67	— 0.42	16 «	1.23	52.05	— 0.12	51.46	+ 0.48
21 «	1.45	49.70	0	49.96	— 0.26	21 «	1.21	50.73	0	50.03	+ 0.70
26 «	1.56	+ 1.21	+ 0.33	51.26	0	26 «	2.13	— 1.63	— 0.50	48.60	0
2 Mai	1.56	52.13	+ 0.69	+ 1.12	+ 0.44	1 Nov.	1.65	47.44	— 0 49	— 1.41	— 0.24
7 «	1.50	53.34	+ 0.98	53.50	+ 0.82	6 «	1.71	45.82	— 0.58	45.77	— 0.53
12 «	0 29	54.56	— 0.01	54.62	— 0.07	11 «	1.04	44.20	0	44.36	+ 0.16
17 «	1.18	55.79	0	55.74	+ 0.05	16 «	0.95	— 0.76	+ 0.19	42.95	+ 0.30
22 «	1.28	+ 0.95	+ 0.34	56.37	+ 0.20	21 «	1.72	42.68	— 1.15	41.53	0
27 «	0.92	57.69	+ 0.30	57.99	0	26 «	0.58	41.92	— 0.97	— 0.70	+ 0.12
1 Juin	0.26	58.64	— 0.39	+ 0.66	— 0.40	1 Déc.	+ 0.20	41.15	0	40.13	+ 0.76
6 «	1.37	59.60	+ 0.02	59.32	+ 0.30	6 «	1.55	— 0.85	— 0.42	39.43	+ 0.27
11 «	0.93	60.55	0	59.98	+ 0.57	11 «	0.91	39.19	— 0.40	38.73	+ 0.06
16 «	0.57	+ 0.55	+ 0.02	60.65	+ 0.47	16 «	0.75	38.33	— 0.29	38.04	0
21 «	0.06	61.65	— 0.47	61.31	— 0.13	21 «	0.64	37.40	0	— 1.04	+ 0.40
26 «	0.77	62.20	— 0.25	61.97	— 0.02	26 «	1.61	— 1.14	— 0.47	35.96	— 0.17
2 Juill.	0.68	62.75	— 0.12	62.63	0	31 «	1.05	35.12	— 0.38	34.92	— 0.18
7 «	0.70	63.31	+ 0.02	+ 0.35	+ 0,35	5 Juin	1.44	33.98	— 0.68	33.88	— 0.58
12 «	0.53	63.86	0	63.33	+ 0.53	10 «	0.72	32.58	0	32.58	0

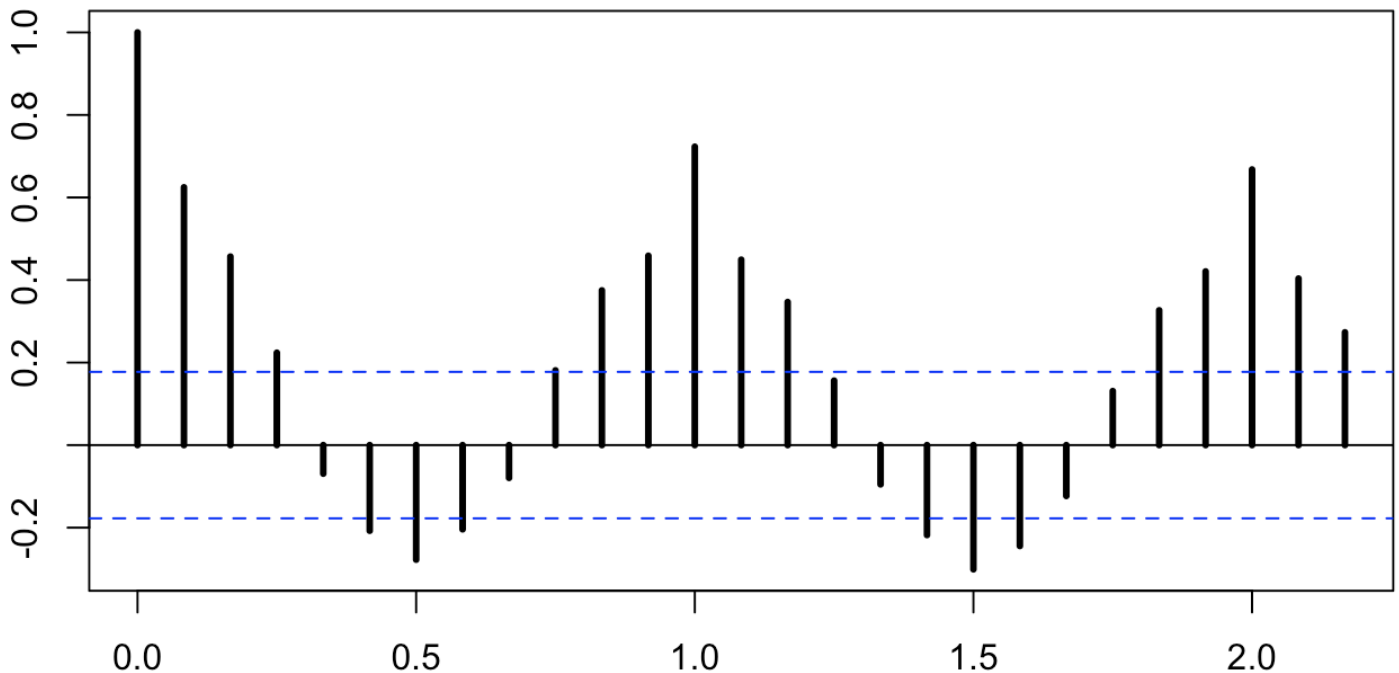
Moving Average

$$\widetilde{y}_t = \frac{1}{2h}\left(\frac{y_{t-h}}{2} + \cdots + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+h}}{2}\right)$$



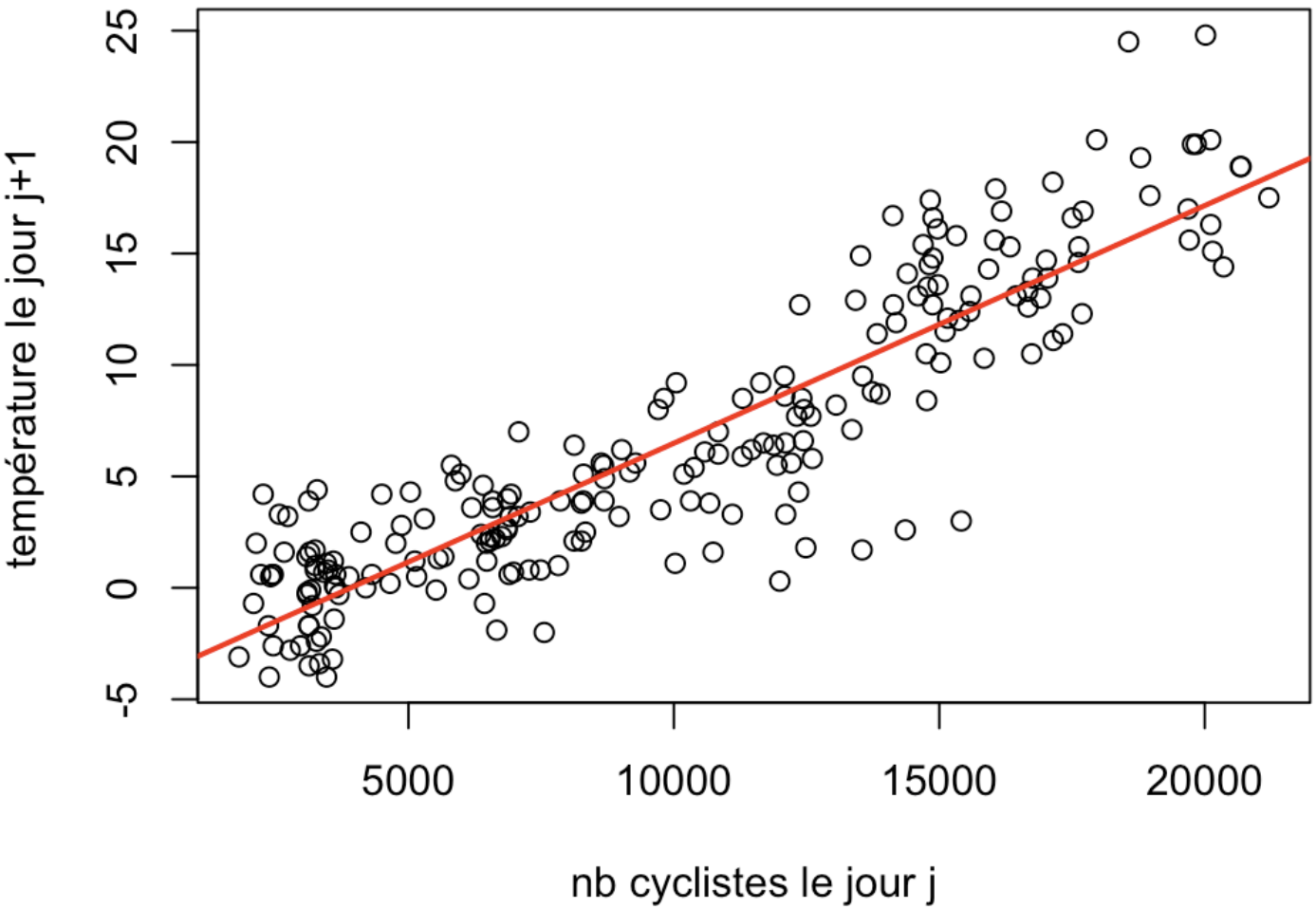
Autocorrelation

$$h \mapsto \text{corr}(y_t, y_{t-h})$$



Time Series & Causality

"*an old variable explains 85% of the change in a new variable. So we can talk about causality*" Emmanuel Todd (no!!)



Time Series & Causality

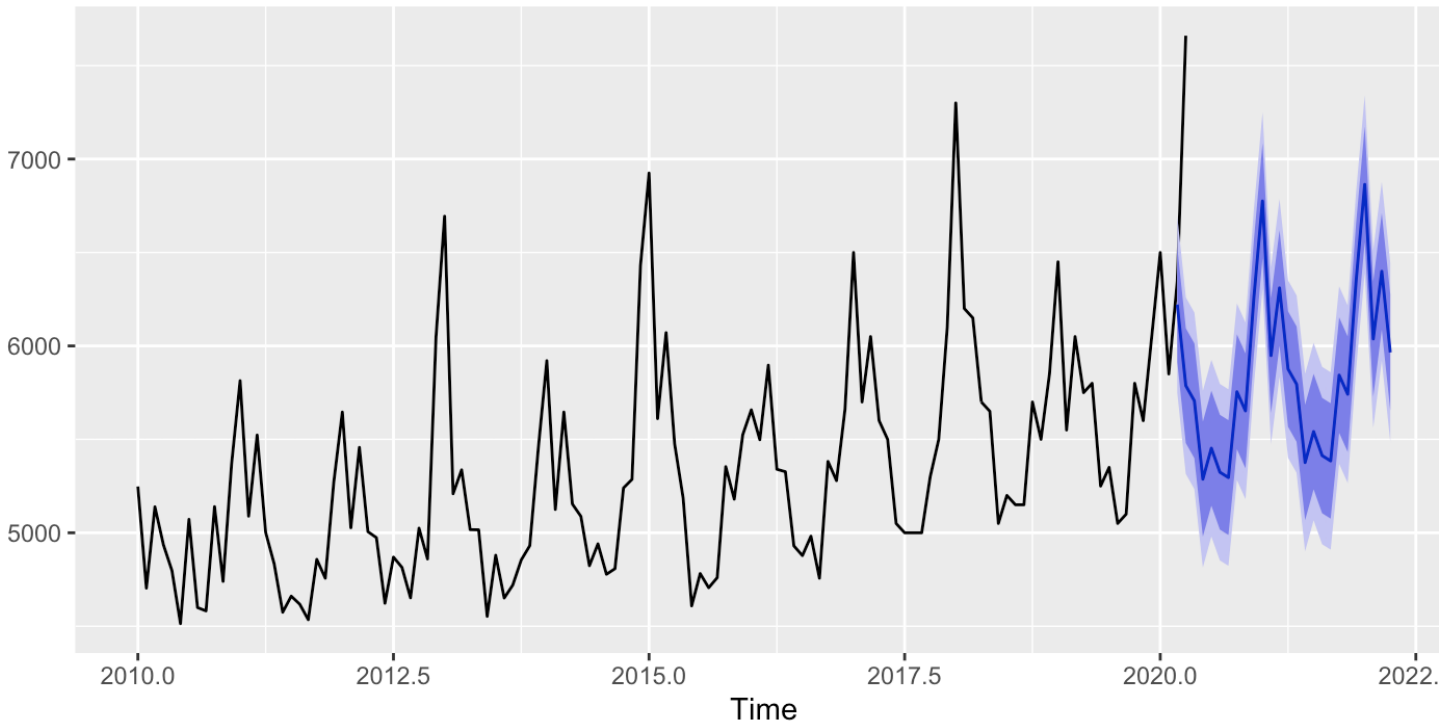
$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}, \ \text{Var} \begin{bmatrix} u_t \\ v_t \end{bmatrix} = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_u \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$$

- $a_{1,2} \neq 0$: causal effect, $y \rightarrow x$
- $a_{2,1} \neq 0$: causal effect, $x \rightarrow y$
- $\rho \neq 0$: (instantaneous) causal effect, $x \leftrightarrow y$

see Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media (on Indignados, Occupy, and Brazilian Vinegar protests)

Forecasting Time Series

See monthly deaths in Québec province



using Exponential Smoothing...