DATA VISUALIZATION COURSE

TIME SERIES

Presenter: Nguyen Thi Kim Nhuong

Instructor: Phuc Loi Luu, PhD

Outline

Part A: Introduction

Part B: Time Series Visualization in R

Part C: How to visualize time series data

What is the best strategy ?

Part D: Forecasting

Part E: Time Series in Biology

2. Types of time series data

1. What is time-series data?

3. Identifying time series data

Part A: Introduction

7. Time series analysis methods

4. Time series data vs cross-sectional and panel data

6. Time series Visualization

5. How is time series data understood and used?

A. Introduction

1. What is time-series data?

Time-series is a ubiquitous type of data set which **describes** how some **measurable feature** (population, snowfall, items sold) has **changed** over a period of time (1).

Time-series **is** a sequence of data in **chronological order** (2).

- Data is **commonly recorded sequential**ly, over time
- Time series data is everywhere (financial of economic, CPI, stock market)

Time series data is a **collection** of observations (**behavior**) for a single subject (**entity**) at **different time intervals** (generally **equally spaced** as in the case of **metrics**, or **unequally spaced** as in the case of **events**) (3).

The relevance of time as an axis makes time series data distinct from other types of data.

A. Introduction

2. Types of time series data

Sampling frequency:

1. exact (temperature measurements for every hour in a day)
2. approximate (temperature measurements recorded every time you check your email)
3. missing values (temperature measurements while you are awake)

Weather records, economic indicators and patient health evolution metrics — all are time series data.

A. Introduction

2. Types of time series data

Example of time series data

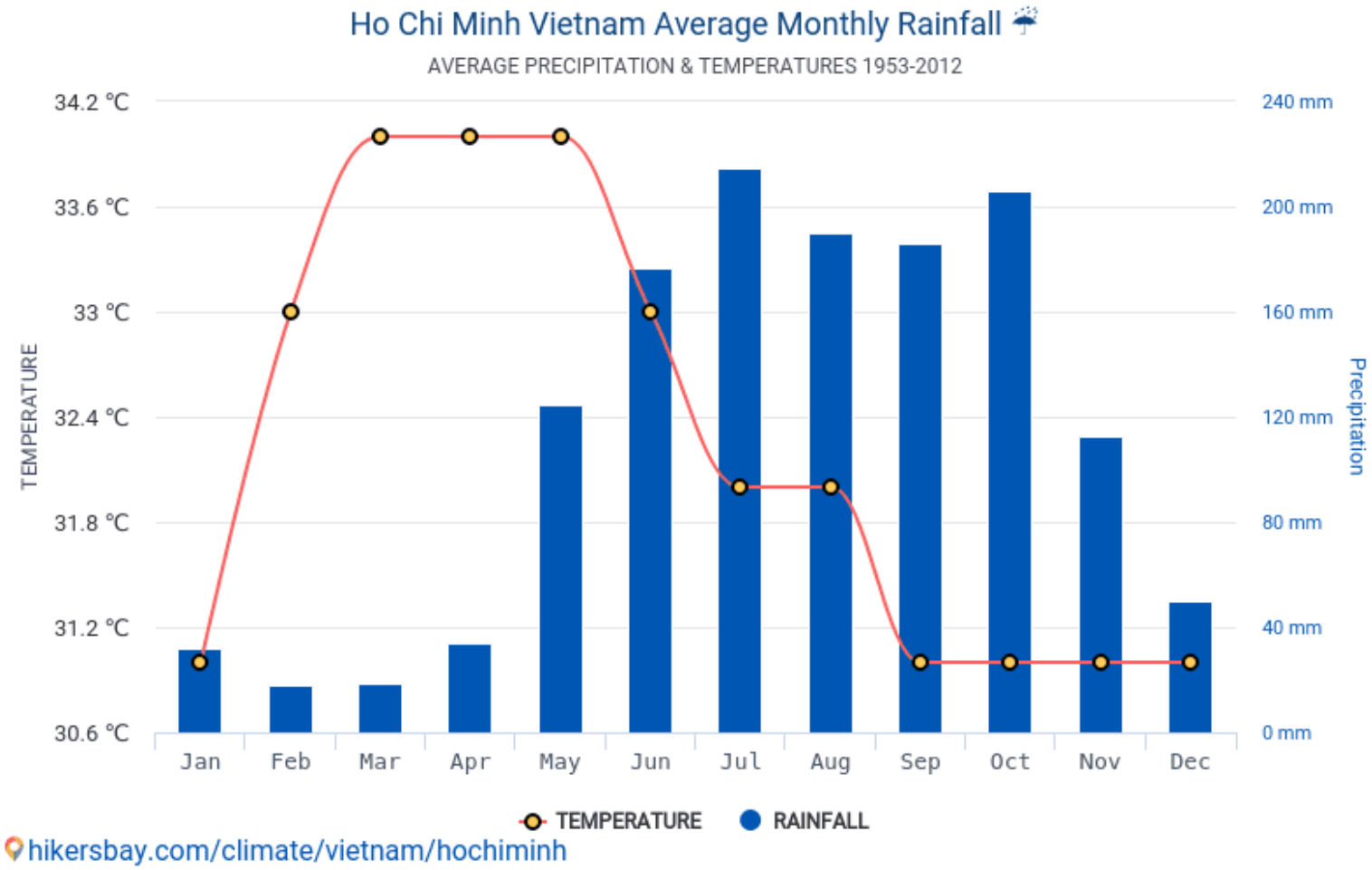
- Rainfall measurements
- Electrical activity in the brain
- Stock prices
- Number of sunspots
- Annual retail sales
- Monthly subscribers
- Heartbeats per minute
- COVID-19 positive cases

A. Introduction

2. Types of time series data

Example of time series data

Rainfall measurements



- A time series plot is a graph
  - The **x-axis** is labeled as the **time-axis.**
  - The **y-axis** represents the **variable being measured.**

A. Introduction

2. Types of time series data

Example of time series data

Stock exchange

Exception

- In investing, a time series **tracks** the **movement** of **data points.**
  - such as a **security's price** over a specified period of time with data points recorded at regular intervals.

- This can be tracked over the short term or the long term

https://www.influxdata.com/what-is-time-series-data/

A. Introduction 2. Types of time series data

Example of time series data

Cluster monitoring

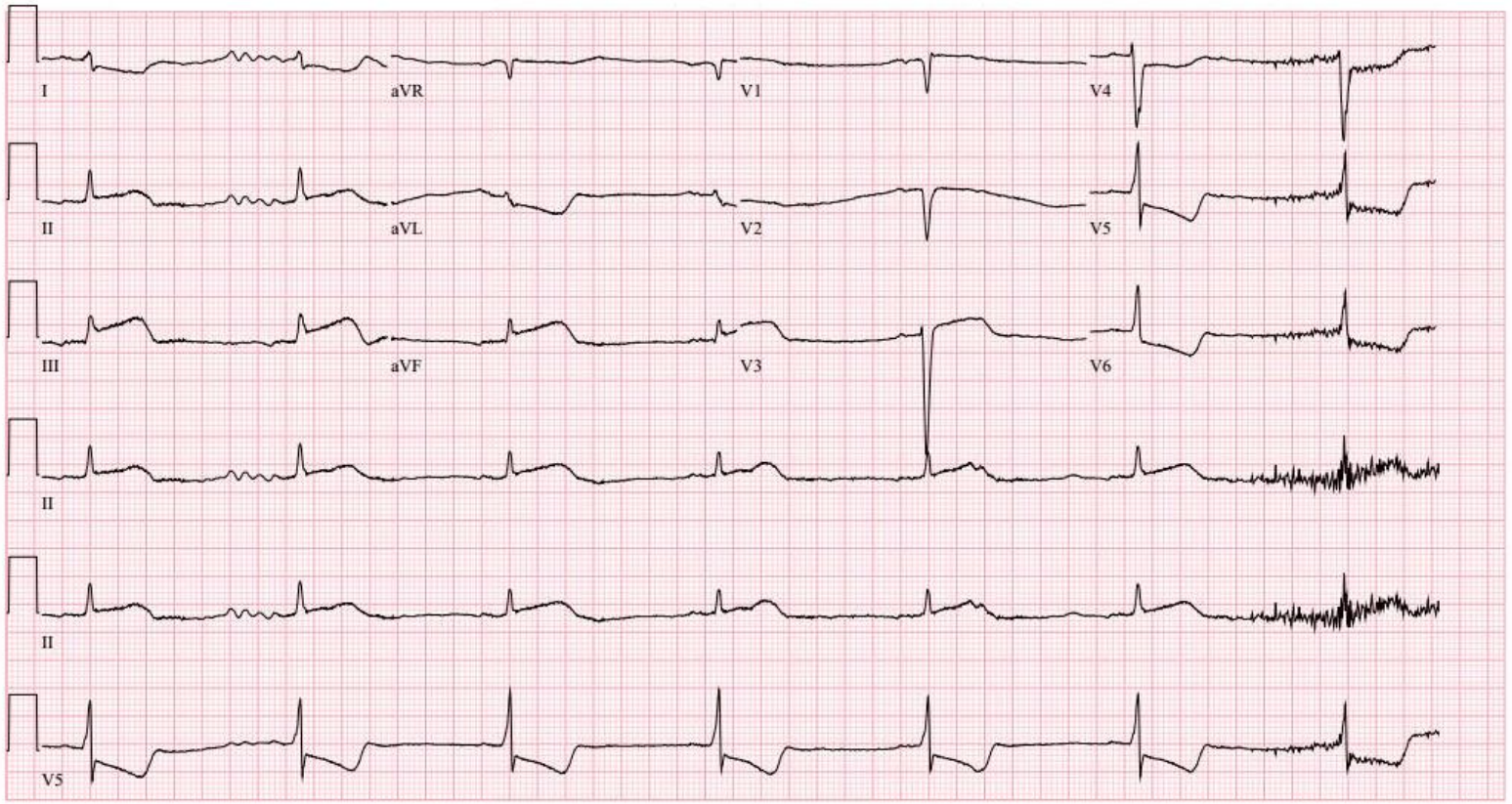
- Depicting disk ops write and usage data, would be familiar to Network Operation Center teams.
- Monitoring data is time series data.

https://www.influxdata.com/what-is-time-series-data/

A. Introduction 2. Types of time series data

Example of time series data

Health monitoring



- Electrocardiogram (ECG), which monitors the heart's activity to show whether it is working normally.

https://www.influxdata.com/what-is-time-series-data/

A. Introduction 2. Types of time series data

Example of time series data

- In addition to being captured at regular time intervals, **time series data** can be **captured whenever it happens** — regardless of the time interval, such as in logs.
- Logs are a registry of events, processes, messages and communication between software applications and the operating system.
- Every executable file produces a log file where all activities are noted. Log data is an important contextual source to triage and resolve issues.

https://www.influxdata.com/what-is-time-series-data/

A. Introduction 2. Types of time series data

Example of time series data

Logs Eg: networking, an event log helps provide information about network traffic, usage & other conditions.

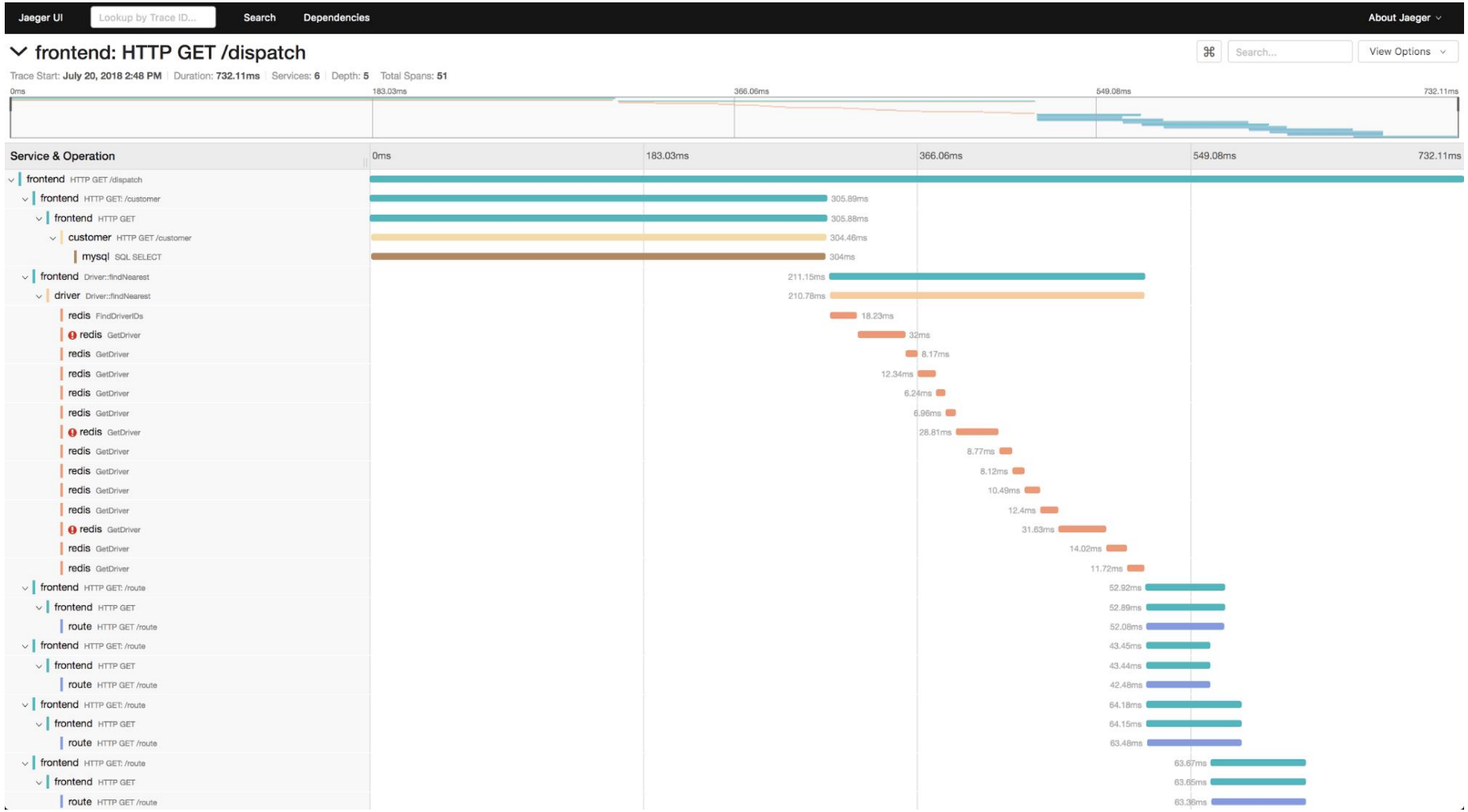
```
Jun 24 13:45:36 haproxy epa-http.txt: dd14-034.compuserve.com [30:01:46:50] "GET /logos/small_gopher.gif HTTP/1.0" 200 935
Jun 24 13:45:38 haproxy epa-http.txt: dd14-034.compuserve.com [30:01:46:54] "GET /logos/small_ftp.gif HTTP/1.0" 200 124
Jun 24 13:45:40 haproxy epa-http.txt: ix-eve-wa2-02.ix.netcom.com [30:01:46:55] "GET /docs/EPA-WASTE/1994/October/Day-05 HTTP/1.0" 302 -
Jun 24 13:45:40 haproxy epa-http.txt: dd14-034.compuserve.com [30:01:46:56] "GET /icons/book.gif HTTP/1.0" 200 156
Jun 24 13:45:41 haproxy epa-http.txt: ix-eve-wa2-02.ix.netcom.com [30:01:46:56] "GET /EPA-WASTE/1994/October/Day-05/ HTTP/1.0" 200 623
Jun 24 13:45:42 haproxy epa-http.txt: dd14-034.compuserve.com [30:01:46:58] "GET /logos/us-flag.gif HTTP/1.0" 200 2788
Jun 24 13:45:43 haproxy epa-http.txt: ix-eve-wa2-02.ix.netcom.com [30:01:47:12] "GET /docs/EPA-WASTE/1994/October/Day-03 HTTP/1.0" 302 -
Jun 24 13:45:45 haproxy epa-http.txt: ix-eve-wa2-02.ix.netcom.com [30:01:47:14] "GET /EPA-WASTE/1994/October/Day-03/ HTTP/1.0" 200 785
Jun 24 13:45:46 haproxy epa-http.txt: dd14-034.compuserve.com [30:01:47:19] "GET /icons/ok2-0.gif HTTP/1.0" 200 231
Jun 24 13:45:48 haproxy epa-http.txt: bettong.client.uq.oz.au [30:01:47:24] "GET /enviro/html/emci/emci_overview.html HTTP/1.0" 200 2352
Jun 24 13:45:49 haproxy epa-http.txt: bettong.client.uq.oz.au [30:01:47:31] "GET /enviro/gif/efacts.gif HTTP/1.0" 200 1367
Jun 24 13:45:50 haproxy epa-http.txt: 202.96.29.111 [30:01:47:34] "GET /PressReleases/ HTTP/1.0" 200 1241
Jun 24 13:45:51 haproxy epa-http.txt: bettong.client.uq.oz.au [30:01:47:37] "GET /enviro/gif/blueball.gif HTTP/1.0" 200 903
Jun 24 13:45:53 haproxy epa-http.txt: ix-eve-wa2-02.ix.netcom.com [30:01:47:37] "GET /Rules.html HTTP/1.0" 200 3273
Jun 24 13:45:53 haproxy epa-http.txt: 202.96.29.111 [30:01:47:38] "GET /icons/circle_logo_small.gif HTTP/1.0" 200 2624
Jun 24 13:45:54 haproxy epa-http.txt: 202.96.29.111 [30:01:48:04] "POST /cgi-
  bin/waisgate/134.67.99.11=earth1.epa.gov=210=/usr1/comwais/indexes/PressReleases=gopher%40earth1=0.00=:free HTTP/1.0" 200 3993
Jun 24 13:45:54 haproxy epa-http.txt: 202.96.29.111 [30:01:48:16] "GET /waisicons/text.xbm HTTP/1.0" 200 527
Jun 24 13:45:55 haproxy epa-http.txt: dd14-034.compuserve.com [30:01:48:22] "GET /Rules.html HTTP/1.0" 200 3273
Jun 24 13:45:57 haproxy epa-http.txt: www-c8.proxy.aol.com [30:01:48:23] "GET /docs/Searchable.html HTTP/1.0" 200 765
Jun 24 13:45:58 haproxy epa-http.txt: bettong.client.uq.oz.au [30:01:48:25] "GET /enviro/gif/banner.gif HTTP/1.0" 200 14887
Jun 24 13:54:14 farm-trivia-72 app/web.1:   User Load (1.2ms)  SELECT  "users".* FROM "users" WHERE "users"."id" = $1  ORDER BY "users"."id" ASC
  LIMIT 1  [["id", 1]]
Jun 24 13:54:14 farm-trivia-72 app/web.1:    (1.3ms)  SELECT COUNT(*) FROM "products"
Jun 24 13:54:14 farm-trivia-72 heroku/router: at=info method=GET path="/a" host=farm-trivia-72.herokuapp.com request_id=3a095914-087a-4b7a-9f88-81d6e2ba7771
  fwd="23.252.53.179" dyno=web.1 connect=1ms service=44ms status=200 bytes=6407
Jun 24 13:54:14 farm-trivia-72 app/web.1:   Product Load (1.4ms)  SELECT  "products".* FROM "products"  ORDER BY products.updated_at desc LIMIT 1
Jun 24 13:54:14 farm-trivia-72 app/web.1:   User Load (1.4ms)  SELECT  "users".* FROM "users"  ORDER BY users.updated_at desc LIMIT 1
Jun 24 13:54:14 farm-trivia-72 app/web.1:    (1.2ms)  SELECT COUNT(*) FROM "users"
Jun 24 13:54:14 farm-trivia-72 app/web.1: method=GET path=/a/ format=html controller=rails_admin/main action=dashboard status=200 duration=35.71 view=20.85 db=6.39
  remote_ip=23.252.53.179 user_id=1 params={}
Jun 24 13:54:16 farm-trivia-72 heroku/router: at=info method=GET path="/a/product?_pjax=%5Bdata-pjax-container%5D" host=farm-trivia-72.herokuapp.com
  request_id=4e7f806e-63b2-493a-88d4-ec8ebab5f0a6 fwd="23.252.53.179" dyno=web.1 connect=3ms service=102ms status=200 bytes=17350
Jun 24 13:54:16 farm-trivia-72 app/web.1:   Product Load (1.7ms)  SELECT  "products".* FROM "products"  ORDER BY products.id desc LIMIT 20 OFFSET 0
Jun 24 13:54:16 farm-trivia-72 app/web.1:   User Load (1.2ms)  SELECT  "users".* FROM "users" WHERE "users"."id" = $1  ORDER BY "users"."id" ASC LIMIT 1
  [["id", 1]]
Jun 24 13:54:16 farm-trivia-72 app/web.1:    (1.3ms)  SELECT COUNT(*) FROM "products"
```

https://www.influxdata.com/what-is-time-series-data/

A. Introduction 2. Types of time series data

Traces

- The **goal** of tracing is to **follow** a program's flow and data progression.
- Tracing encompasses a wide, continuous view of an application **to find bugs in a program or application.**

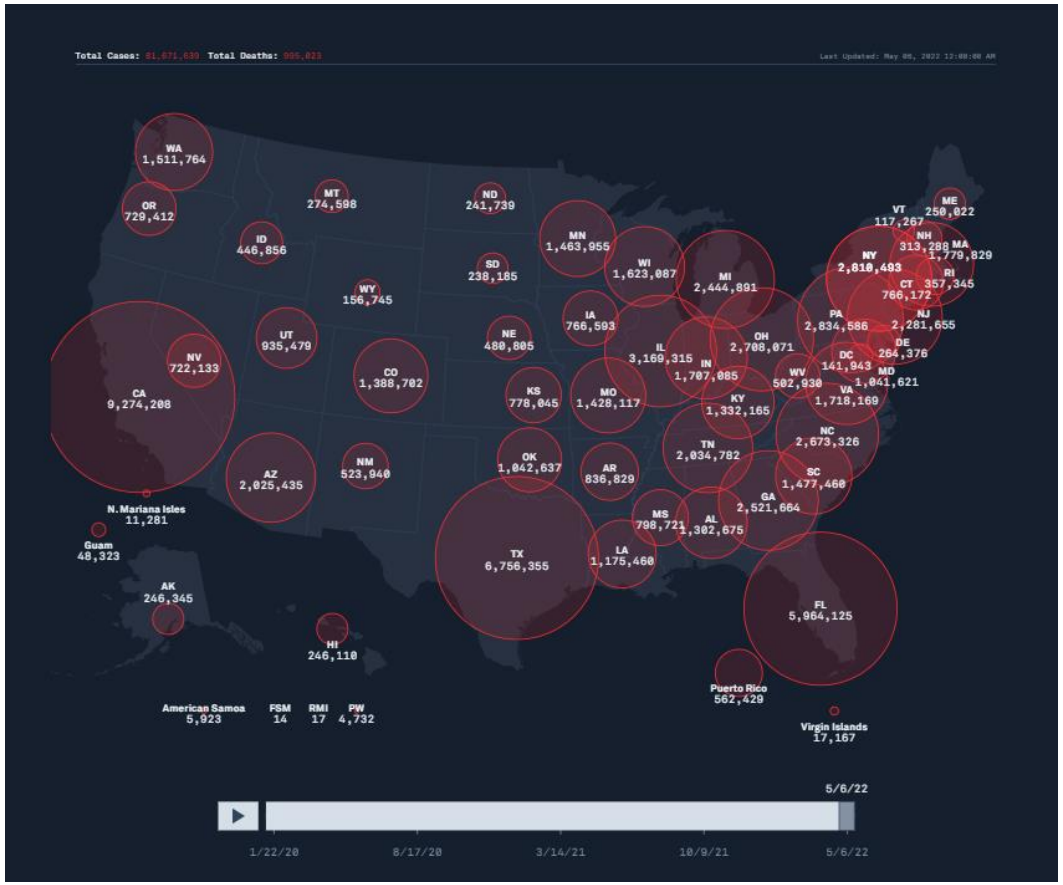


Example Of time series data

https://www.influxdata.com/what-is-time-series-data/

A. Introduction 2. Types of time series data

Example of time series data



Date: 5/6/2022

State	Cases	Deaths	Weekly* Growth
AL	1,302,675	19,601	0.1%
AK	246,345	1,220	0.6%
AZ	2,025,435	30,189	0.1%
AR	836,829	11,404	0.1%
CA	9,274,208	89,851	0.5%
CO	1,388,702	12,481	0.4%
CT	766,172	10,874	0.9%
DE	264,376	2,928	0.6%
DC	141,943	1,340	0.2%
FL	5,964,125	74,085	0.5%
GA	2,521,664	37,969	0.2%
GU	48,323	355	0.4%
HI	246,110	1,416	1.5%
ID	446,856	4,928	0.1%
IL	3,169,315	37,956	0.9%
IN	1,707,085	23,614	0.3%
IA	766,593	9,533	0.3%
KS	778,045	8,691	0.3%
KY	1,332,165	15,700	0.3%

State	Cases	Deaths	Weekly* Growth
LA	1,175,460	17,276	0.2%
ME	250,022	2,325	1.9%
MD	1,041,621	14,495	0.7%
MA	1,779,829	20,334	1.2%
MI	2,444,891	36,064	0.7%
MN	1,463,955	12,805	0.7%
MS	798,721	12,454	0.1%
MO	1,428,117	20,382	0.3%
MT	274,598	3,378	0.2%
NE	480,805	4,216	0.2%
NV	722,133	10,787	0.3%
NH	313,288	2,487	0.9%
NJ	2,281,655	33,482	0.9%
NM	523,940	7,556	0.3%
NY	2,810,493	27,554	1.2%
NC	2,673,326	24,577	0.4%
ND	241,739	2,268	0.2%
OH	2,708,071	38,495	0.3%
OK	1,042,637	14,346	0.2%

State	Cases	Deaths	Weekly* Growth
OR	729,412	7,528	1.1%
PA	2,834,586	44,757	0.5%
PR	562,429	4,227	5.2%
RI	357,345	3,544	1.0%
SC	1,477,460	17,847	0.2%
SD	238,185	2,915	0.1%
TN	2,034,782	28,208	0.2%
TX	6,756,355	86,619	0.3%
UT	935,479	4,760	0.2%
VT	117,267	611	1.9%
VI	17,167	113	2.8%
VA	1,718,169	20,288	0.8%
WA	1,511,764	12,731	1.0%
WV	502,930	6,880	0.3%
WI	1,623,087	14,464	0.6%
WY	156,745	1,814	0.1%
AS	5,923	28	1.6%
MP	11,281	34	0.2%

* Here's how we calculated the nation's growth rate of COVID-19 cases over the last week. Take today's seven-day moving average of confirmed cases. Divide that number by the seven-day moving average of confirmed cases from a week ago and subtract one. Then, multiply by 100.

Watch the real-time spread of coronavirus in the U.S.

https://www.pbs.org/newshour/features/coronavirus/us/

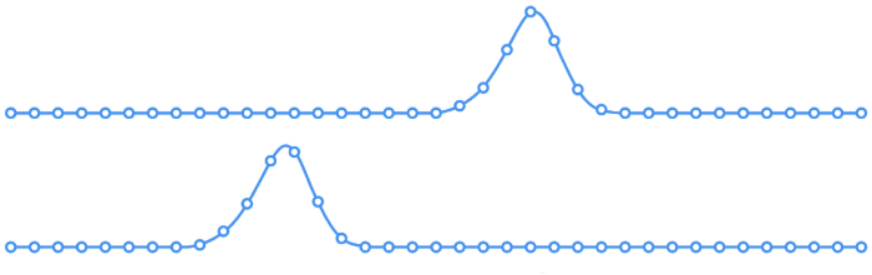
A. Introduction

2. Types of time series data

How many type of time series data ?

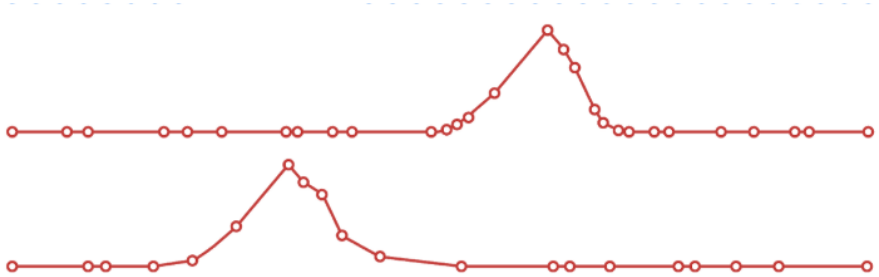
① **Metrics (Regular)**

Measurements gathered at **regular time intervals**



② **Event (Irregular)**

Measurements gathered at **irregular time intervals**

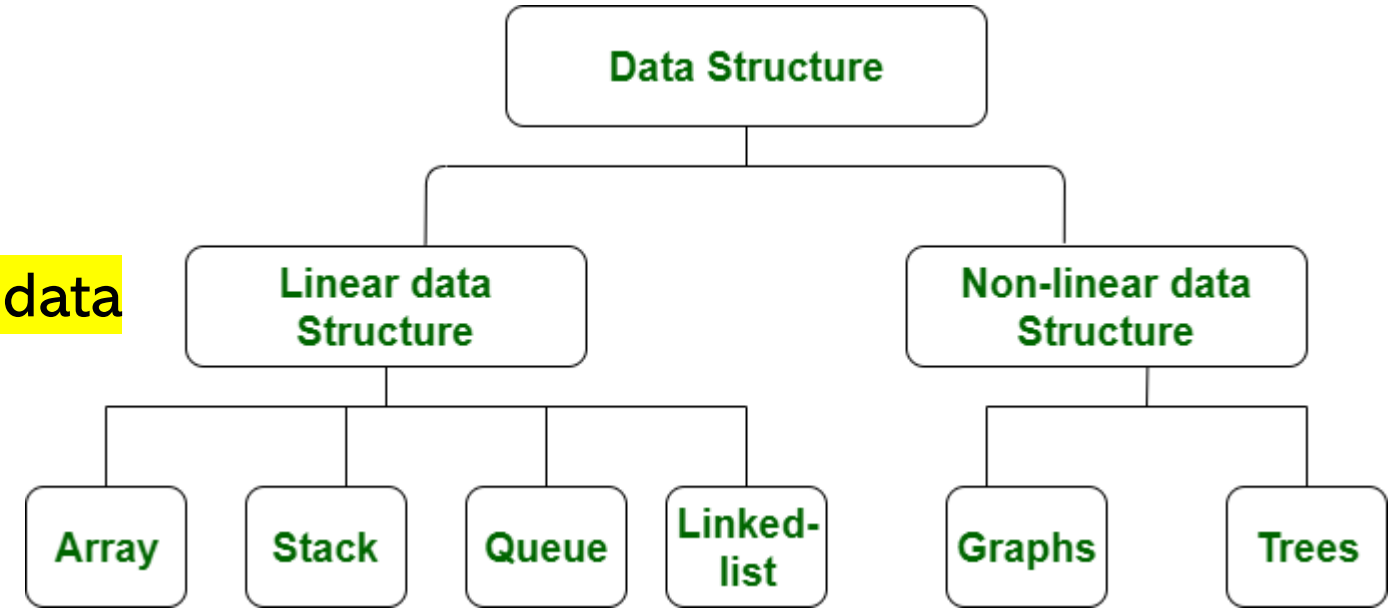


Can we predict event time series data ?

A. Introduction

2. Types of time series data

Linear vs. nonlinear time series data



- A linear time series is one where, for **each data point Xt**, that data point can be viewed as a **linear combination** of past or future values or differences.
- Nonlinear time series are generated by **nonlinear dynamic equations.** They have features that cannot be **modelled** by **linear processes:** time-changing variance, asymmetric cycles, higher-moment structures, thresholds and breaks.

A. Introduction

2. Types of time series data

Linear vs. nonlinear time series data

- Here are some important considerations when working with linear and nonlinear time series data:

1. If a regression equation doesn't follow the rules for a linear model, then it must be a nonlinear model.
2. Nonlinear regression can fit an enormous variety of curves.
3. The defining characteristic for both types of models are the functional forms.

A. Introduction

2. Types of time series data

Different forms of time series data

- Time series data is not always numeric
- It can be int64, float64, bool, or string.

A. Introduction

3. Identifying time series data

Immutability

- Since time series data **comes in time order**, it is almost always **recorded in a new entry**, and as such, should be immutable and **append-only (appended to the existing data).**
- The fact that time series data is ordered makes it unique in the data space because it often displays serial dependence. "Autocorrelation in Time Series Data".
- No events that exist outside of time
- Time series data sometimes **exists at high levels** of **granularity**, as frequently as **microseconds** or even **nanoseconds**.

A. Introduction

3. Identifying time series data

How to define whether working data is time series data ?

- ❑ If all you **need** is a **timestamp**, it's probably **time series data.**
- ❑ If you **need something other than** a timestamp, it's probably **cross-sectional data.**
- ❑ If you **need a timestamp plus something** else, like an ID, it's probably **panel data.**

A. Introduction

4. Time series data vs cross-sectional and panel data

Is this a time series data ?

CASE STUY: Max Temperature, Humidity and Wind (all three behaviors) in New York City, SFO, Boston, Chicago (**multiple entities**) on 1/1/2015 (**single instance**).

Cross-sectional data definition

- Cross-sectional data is a collection of **observations** (behavior) for **multiple subjects** (entities such as different individuals or groups) at **a single point in time.**
- For example: In cross-sectional studies, there is **no natural ordering** of the **observations** (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order).

A. Introduction

4. Time series data vs cross-sectional and panel data

Differences between the three data types

- A time series is a **group of observations** on **a single entity over time.**
- A cross-section is a **group of observations** of **multiple entities** at **a single time.**
- If your data is organized in both then you have panel data.

A. Introduction

5. How is time series data understood and used?

Purposes

1. In **data mining**, pattern recognition and machine learning, time series analysis is used for **clustering**, **classification**, **query** by content, anomaly **detection** and **forecasting**.

2. In **signal processing**, control engineering and communication engineering, time series data is used for **signal detection** and **estimation**.

3. In **statistics, econometrics, quantitative finance, seismology, meteorology, and geophysics** the time series analysis is used for forecasting.

A. Introduction

5. How is time series data understood and used?

Why do people need to visualize time series data

#install.packages ("gapminder")

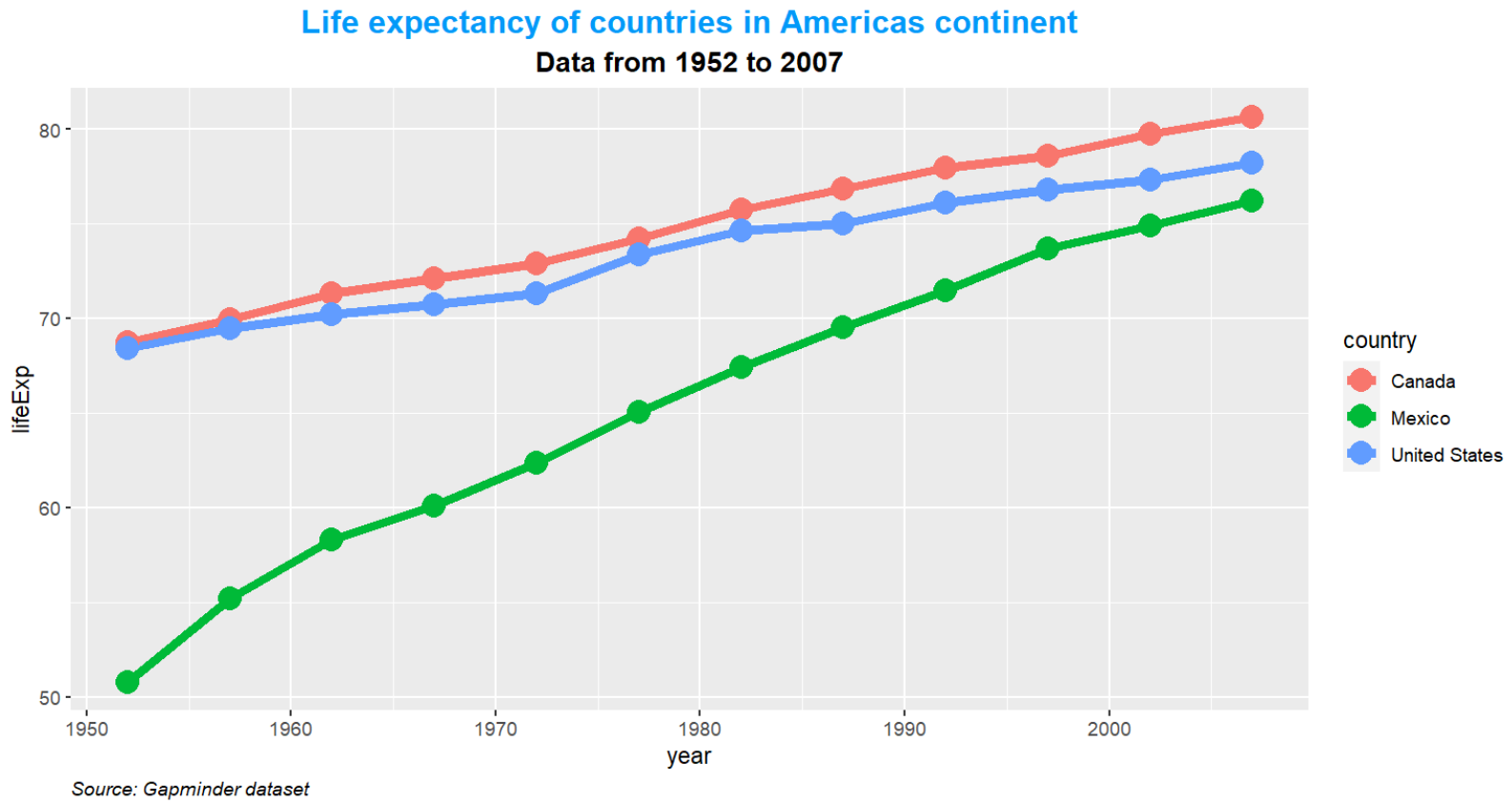
	country	continent	year	lifeExp	pop	gdpPercap
	<fct>	<fct>	<int>	<dbl>	<int>	<dbl>
1	Afghanistan	Asia	1952	28.8	8425333	779.
2	Afghanistan	Asia	1957	30.3	9240934	821.
3	Afghanistan	Asia	1962	32.0	10267083	853.
4	Afghanistan	Asia	1967	34.0	11537966	836.
5	Afghanistan	Asia	1972	36.1	13079460	740.
6	Afghanistan	Asia	1977	38.4	14880372	786.
7	Afghanistan	Asia	1982	39.9	12881816	978.
8	Afghanistan	Asia	1987	40.8	13867957	852.
9	Afghanistan	Asia	1992	41.7	16317921	649.
10	Afghanistan	Asia	1997	41.8	22227415	635.

A. Introduction

5. How is time series data understood and used?

Why do people need to visualize time series data

Compare life expectancy of countries in Americas continent from 1950 to 2000



Source: Gapminder dataset

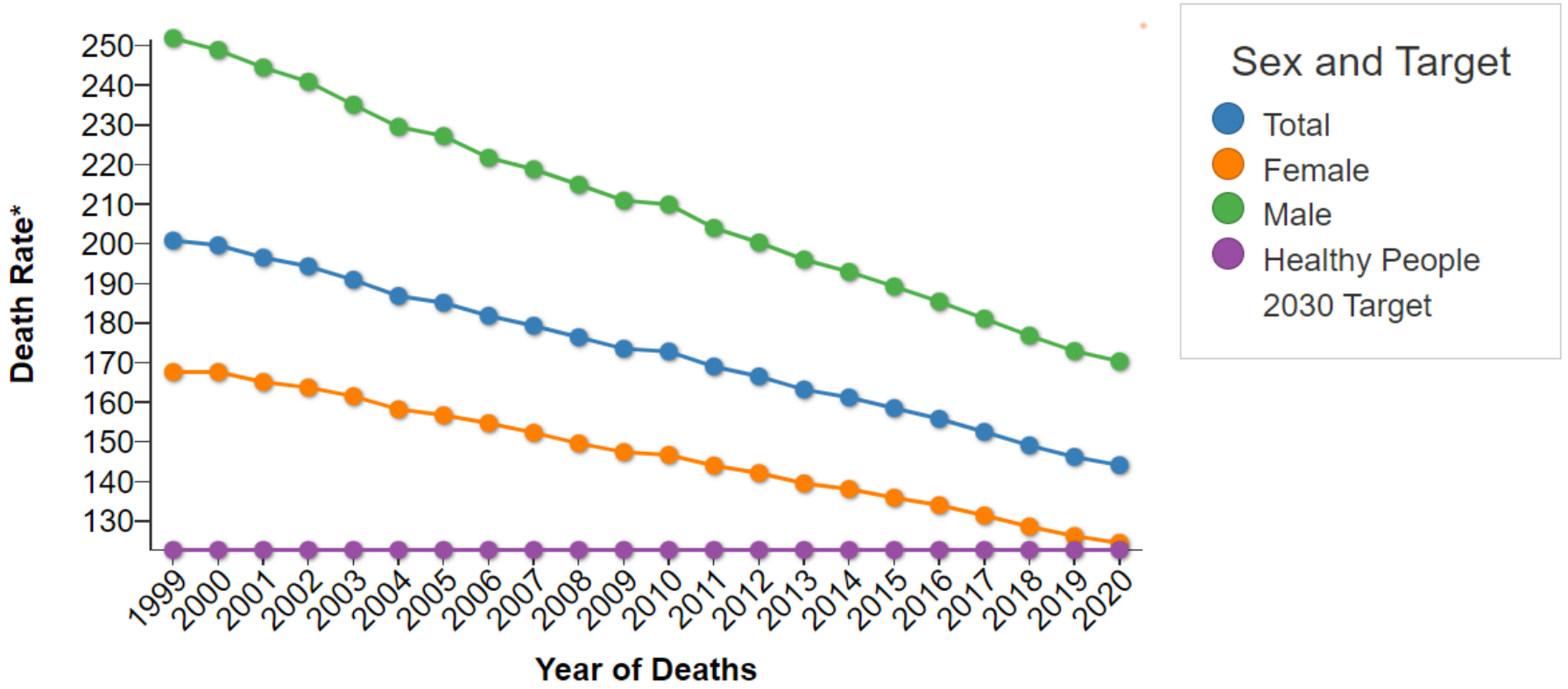
A. Introduction

5. How is time series data understood and used?

Why do people need to visualize time series data

Is cancer increasing or decreasing?

Figure 1. Age-adjusted cancer death rates, by sex, United States, 1999–2020



A. Introduction

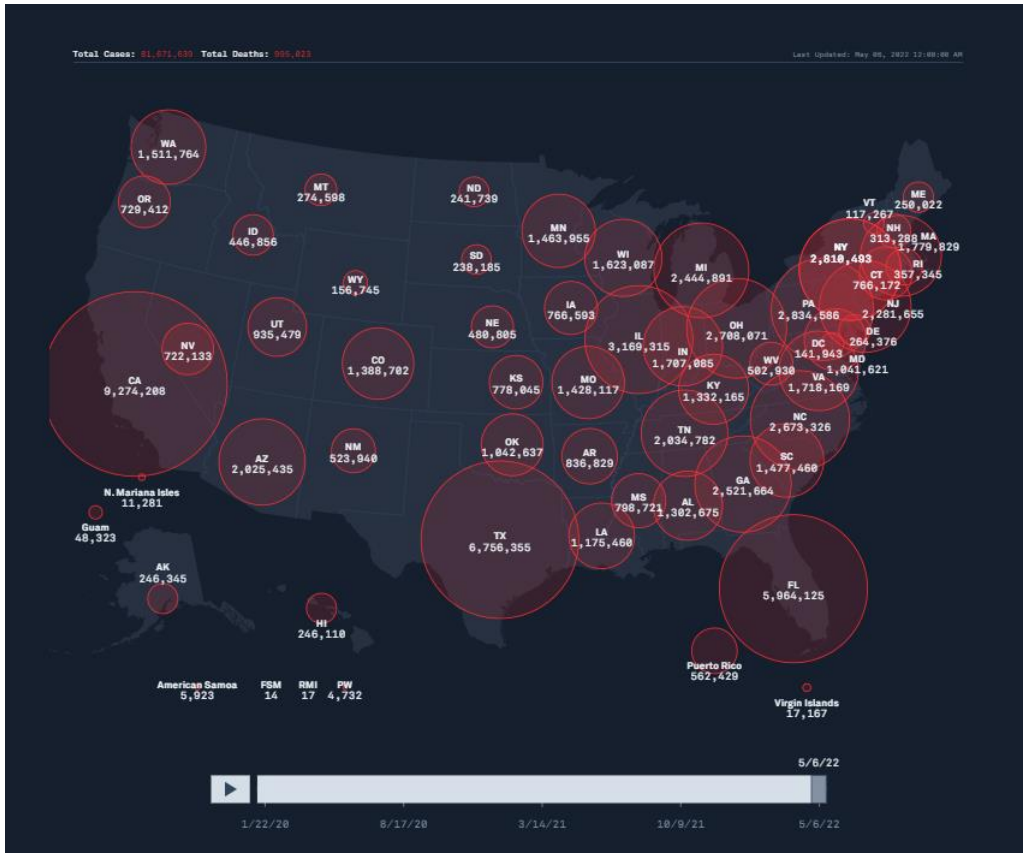
5. How is time series data understood and used?

Why do people need to visualize time series data

Date: 5/6/2022

State	Cases	Deaths	Weekly Growth*	State	Cases	Deaths	Weekly Growth*	State	Cases	Deaths	Weekly Growth*
AL	1,302,675	19,601	0.1%	LA	1,175,460	17,276	0.2%	OR	729,412	7,528	1.1%
AK	246,345	1,220	0.6%	ME	250,022	2,325	1.9%	PA	2,834,586	44,757	0.5%
AZ	2,025,435	30,189	0.1%	MD	1,041,621	14,495	0.7%	PR	562,429	4,227	5.2%
AR	836,829	11,404	0.1%	MA	1,779,829	20,334	1.2%	RI	357,345	3,544	1.0%
CA	9,274,208	89,851	0.5%	MI	2,444,891	36,064	0.7%	SC	1,477,460	17,847	0.2%
CO	1,388,702	12,481	0.4%	MN	1,463,955	12,805	0.7%	SD	238,185	2,915	0.1%
CT	766,172	10,874	0.9%	MS	798,721	12,454	0.1%	TN	2,034,782	26,208	0.2%
DE	264,376	2,928	0.6%	MO	1,428,117	20,382	0.3%	TX	6,756,355	86,619	0.3%
DC	141,943	1,340	0.2%	MT	274,598	3,378	0.2%	UT	935,479	4,760	0.2%
FL	5,964,125	74,085	0.5%	NE	480,805	4,216	0.2%	VT	117,267	611	1.9%
GA	2,521,664	37,969	0.2%	NV	722,133	10,787	0.3%	VI	17,167	113	2.8%
GU	48,323	355	0.4%	NH	313,288	2,487	0.9%	VA	1,718,169	20,288	0.8%
HI	246,110	1,416	1.5%	NJ	2,281,655	33,482	0.9%	WA	1,511,764	12,731	1.0%
ID	446,856	4,928	0.1%	NM	523,940	7,556	0.3%	WV	502,930	6,880	0.3%
IL	3,169,315	37,956	0.9%	NY	2,810,493	27,554	1.2%	WI	1,623,087	14,484	0.6%
IN	1,707,085	23,614	0.3%	NC	2,673,326	24,577	0.4%	WY	156,745	1,814	0.1%
IA	766,593	9,533	0.3%	ND	241,739	2,268	0.2%	AS	5,923	28	1.6%
KS	778,045	8,691	0.3%	OH	2,708,071	38,495	0.3%	MP	11,281	34	0.2%
KY	1,332,165	15,700	0.3%	OK	1,042,637	14,346	0.2%				

* Here's how we calculated the nation's growth rate of COVID-19 cases over the last week. Take today's seven-day moving average of confirmed cases. Divide that number by the seven-day moving average of confirmed cases from a week ago and subtract one. Then, multiply by 100.



A. Introduction

6. Time series Visualization

- Time series data can be **visualized** in **different types** of charts to **facilitate insight extraction**, **trend analysis**, and **anomaly detection.**
- Time series visualization and **dashboarding tools** include the InfluxDB UI and Grafana.
- The term **'time series patterns'** describes **long-term changes** in the **series**.
- Whether **measured** as a **trend**, **seasonal**, or **cyclic pattern**, the **correlation** can be **calculated** in a number of ways (**linear, exponential**, etc.), and the **direction** may **change** at **any given time**.
- Time series data is used in **time series analysis (historical or real-time)** and time series forecasting to detect and predict patterns — essentially looking at change over time.

A. Introduction

7. Time series analysis methods

- **Programming languages** used for analyzing time series.
- **Patterns** that may be present within time series data:
  1. Trend
  2. Seasonality
  3. Residuals
- **Methods** of analyzing time series data:
  1. Frequency-domain methods
  2. Time-domain methods
  3. Others: Parametric & Non-parametric
- **Wavelets** are **analysis tools** mainly for time **series analysis** and **image analysis** . As a subject, wavelets are relatively new (1983 to present) and synthesize many new/old ideas.

A. Introduction

7. Time series analysis methods

Time series analysis best practices

- ✓ understanding of exactly what you're trying to do in the first place.
- ✓ in a time series, the independent variable is often time itself and you're typically using it to try to predict what the future might hold.
- ✓ understanding **whether or not** time is **stationary**, if there is **seasonality**, and if the variable is **autocorrelated**.
  - ➢ Autocorrelation is defined as the **similarity of observations** as a function of the amount of time that passes between them.
  - ➢ Seasonality takes a look at specific, **periodic fluctuations.**
  - ➢ If a time series is stationary, its own **statistical properties** do **not change over time**.

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R & Managing Data/ Time formats

2. Time Series Analysis in R

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R & Managing Data/ Time formats

Data

NEON TEACHING DATA SUBSET: METEOROLOGICAL DATA FOR HARVARD FOREST

https://ndownloader.figshare.com/files/3701572

- collected at the National Ecological Observatory Network's Harvard Forest field site.
- are proxy data for what will be available for 30 years on the NEON data portal for the Harvard Forest and other field sites located across the United States.

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

Data: Holistic View

Data Related to Phenology

- To explore **atmospheric data** (including temperature, precipitation and other metrics)
- Collected by sensors mounted on a flux tower at the NEON Harvard Forest field site.
- To **explore changes** in temperature, precipitation, Photosynthetically Active Radiation (PAR) and day length throughout the year -- metrics that **impact changes** in the timing of plant phenophases (**phenology**).

B. Time Series in R

Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

Import Data

```
# Load csv file of daily meteorological data from Harvard Forest
harMet.daily <- read.csv(
      file=paste0(wd,"NEON-DS-Met-Time-Series/HARV/FisherTower-Met/hf001-06-daily-
m.csv"),
      stringsAsFactors = FALSE
      )
```

stringsAsFactors=FALSE ?

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

Data.Frames in R

How to know what type of R object is our imported data?

Syntax: Class(Name)

The read.csv() imports our .csv into a data.frame object in R. data.frames are ideal for working with tabular data - they are similar to a spreadsheet.

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

Data Structure

head(): shows us the first 6 rows of the data (tail() shows the last 6 rows).

str() : displays the structure of the data as R interprets it.

Classes in R

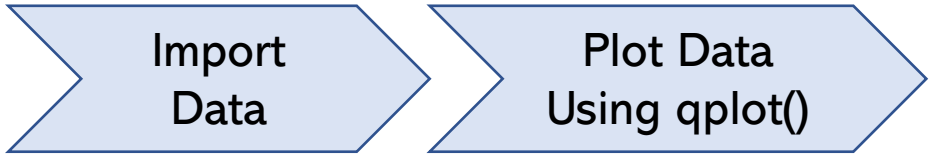
chr – Character

int – Integer

num - Numeric

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats



How to plot air temperature in harMet.dailay data ?

```
# quickly plot air temperature
qplot(x=date, y=airt,
      data=harMet.daily,
      main="Daily Air Temperature\nNEON Harvard Forest Field Site")
```

B. Time Series in R

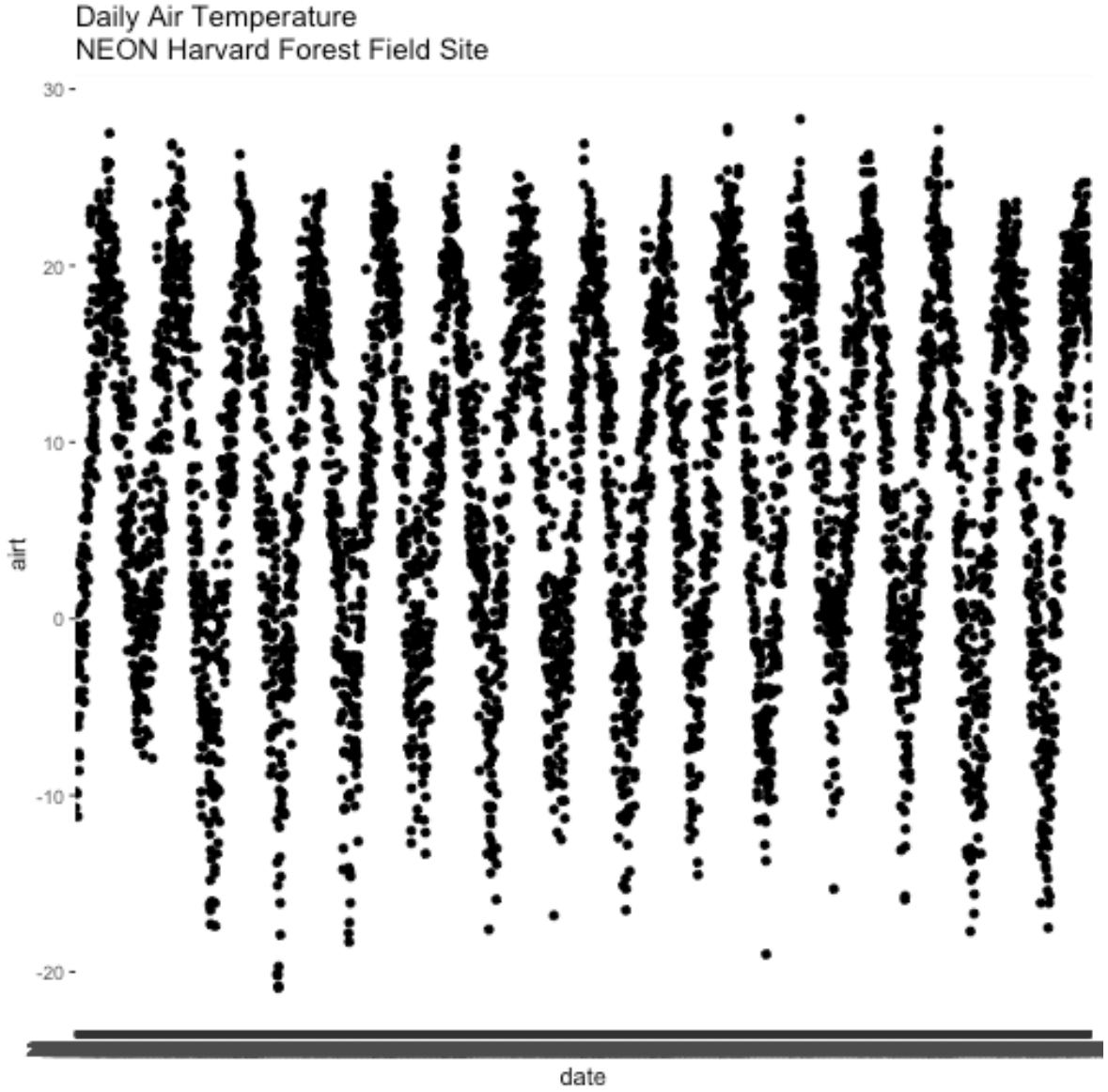
1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

Import Data → Plot Data Using qplot()

Is it easy to view ?

Let's have a look at the class of the x-axis variable - date.

```
# View data class for each column that we wish to plot
class(harMet.daily$date)

## [1] "character"

class(harMet.daily$airt)

## [1] "numeric"
```



B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

Date as a Date-Time Class

```
# convert column to date class
harMet.daily$date <- as.Date(harMet.daily$date)

# view R class of data
class(harMet.daily$date)

## [1] "Date"

# view results
head(harMet.daily$date)

## [1] "2001-02-11" "2001-02-12" "2001-02-13" "2001-02-14" "2001-02-15"
## [6] "2001-02-16"
```

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

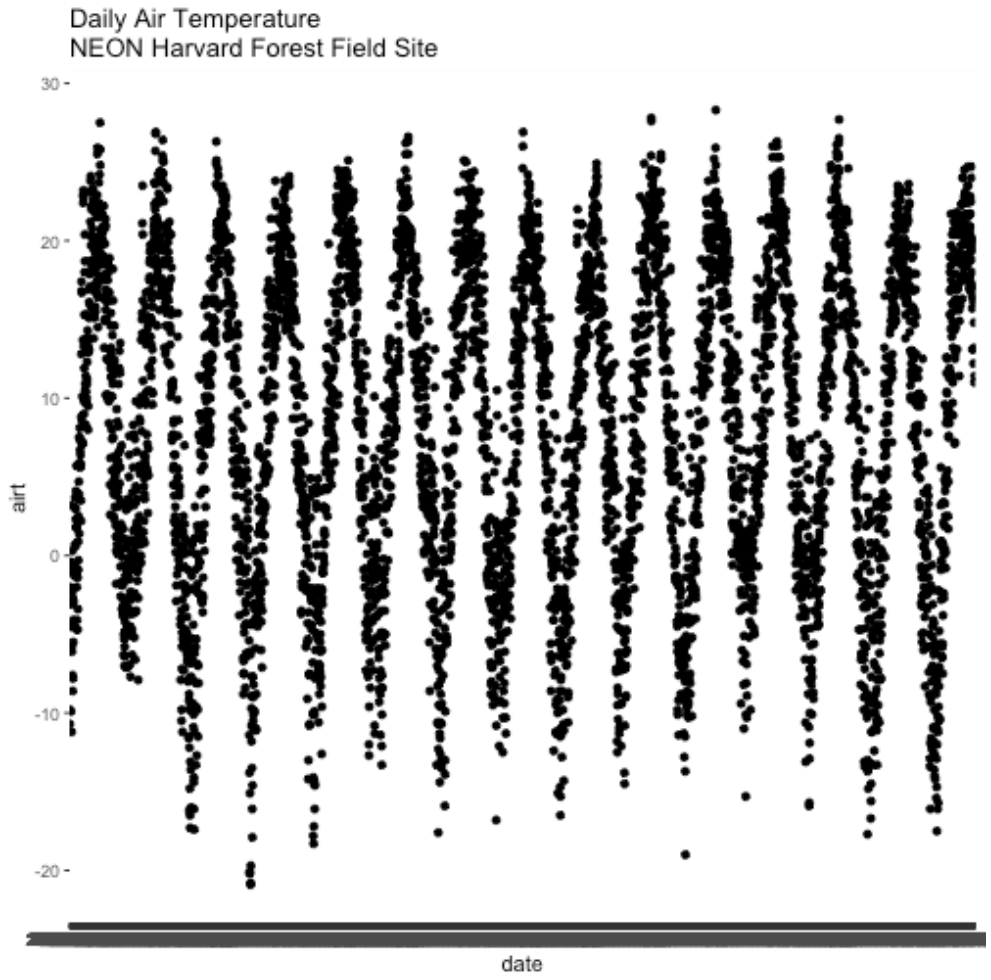
Date as a Date-Time Class

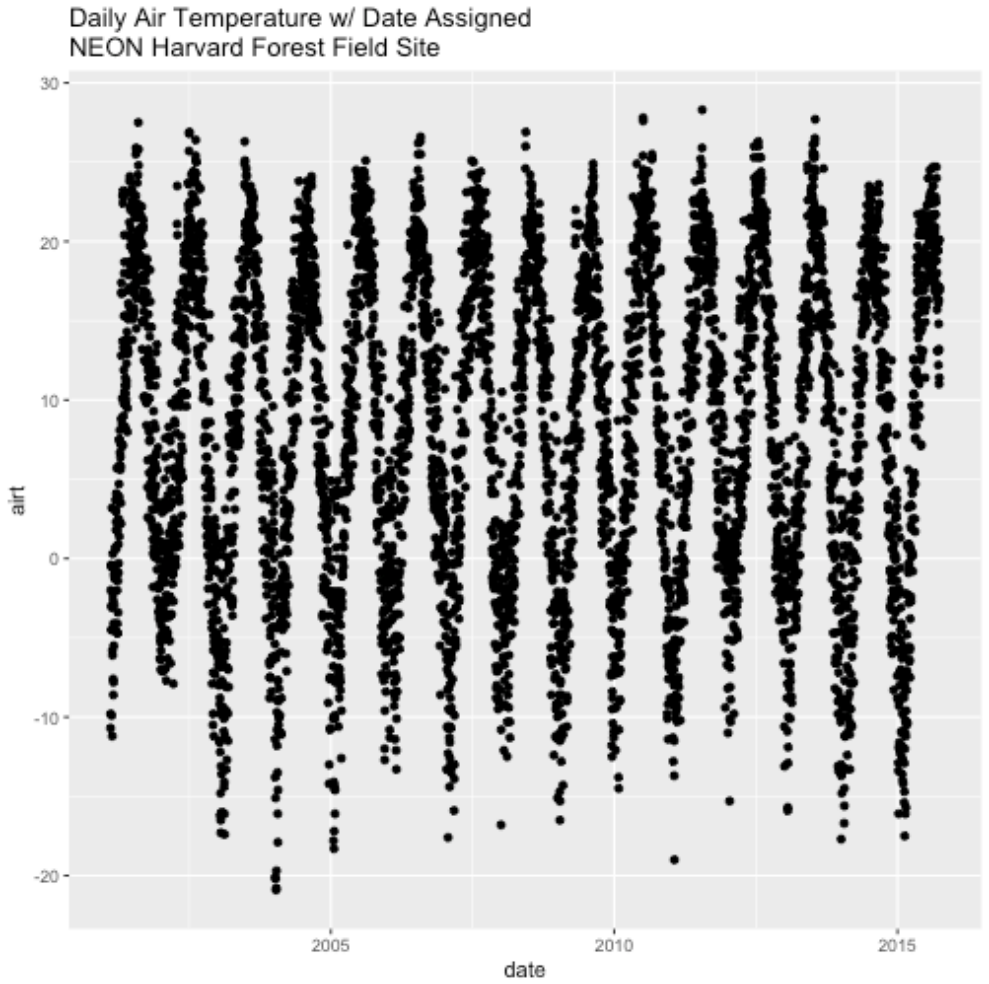
```
# quickly plot the data and include a title using main=""
# In title string we can use '\n' to force the string to break onto a new line
qplot(x=date,y=airt,
      data=harMet.daily,
      main="Daily Air Temperature w/ Date Assigned\nNEON Harvard Forest Field Sit
e")
```

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

Date as a Date-Time Class





B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats

Date as a Date-Time Class

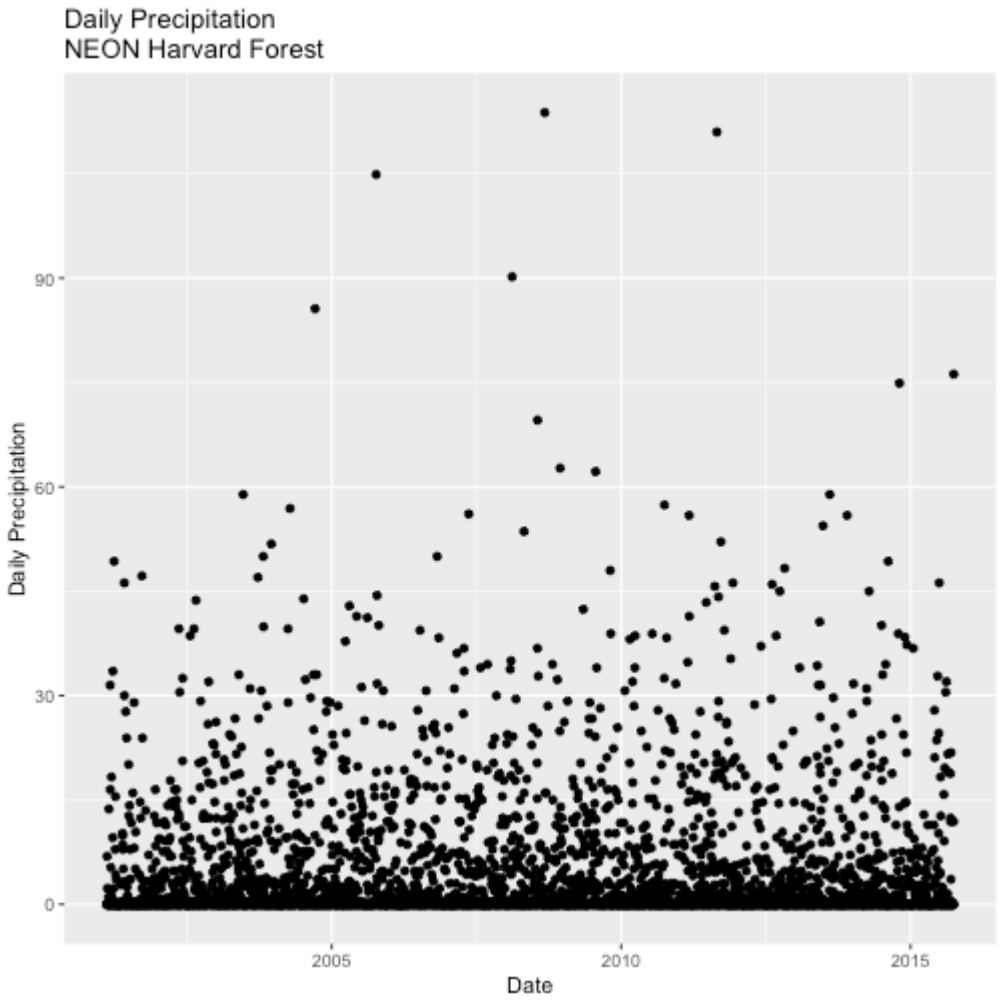
Challenge: Using ggplot2's qplot function

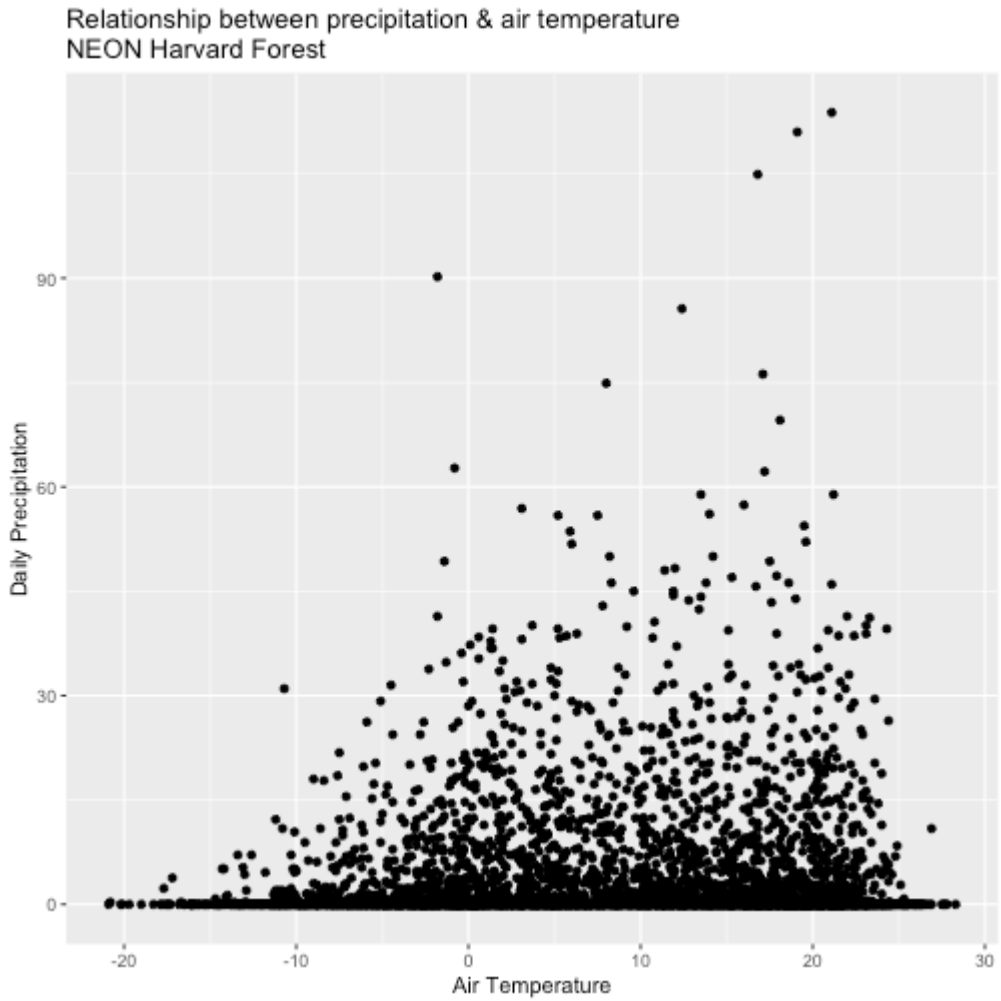
1. Create a quick plot of the precipitation. Use the full time frame of data available in the `harMet.daily` object.
2. Do precipitation and air temperature have similar annual patterns?
3. Create a quick plot examining the relationship between air temperature and precipitation.

Hint: you can modify the X and Y axis labels using `xlab="label text"` and `ylab="label text"`.

B. Time Series in R

1. Tutorial: Intro to Time Series Data in R &Managing Data/ Time formats





B. Time Series in R

2. Time Series Analysis in R

- Time Series in R is used to see how an object behaves over a period of time.
- In R, it can be easily done by ts() function with some parameters.

Syntax: objectName <- ts(data, start, end, frequency)

where,

- data represents the data vector
- start represents the first observation in time series
- end represents the last observation in time series
- frequency represents number of observations per unit time.

  For example, frequency=1 for monthly data.

Example:

- Let's take the example of COVID-19 pandemic situation.
- Taking a total number of positive cases of COVID-19 cases weekly from 22 January 2020 to 15 April 2020 of the world in data vector.

B. Time Series in R

2. Time Series Analysis in R

```
# Weekly data of COVID-19 positive cases from
# 22 January, 2020 to 15 April, 2020
x <- c(580, 7813, 28266, 59287, 75700,
       87820, 95314, 126214, 218843, 471497,
       936851, 1508725, 2072113)
```

```
# library required for decimal_date() function

library(lubridate)
```

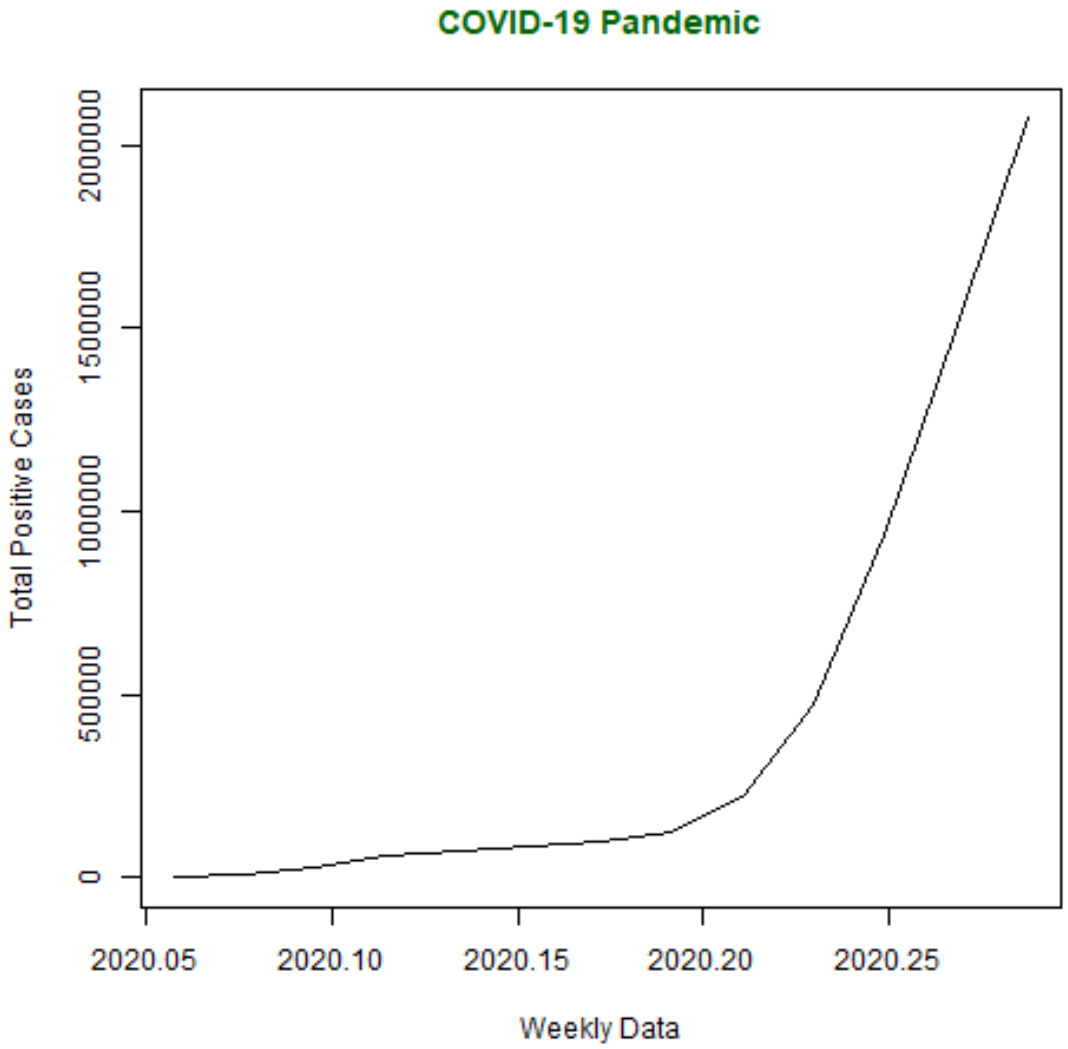
```
# output to be created as png file
png(file ="timeSeries.png")
```

B. Time Series in R

2. Time Series Analysis in R

```
# creating time series object
# from date 22 January, 2020
mts <- ts(x, start = decimal_date(ymd("2020-01-22")),
          frequency = 365.25 / 7)
```

```
# plotting the graph
plot(mts, xlab ="Weekly Data",
     ylab ="Total Positive Cases",
     main ="COVID-19 Pandemic",
     col.main ="darkgreen")

# saving the file
dev.off()
```



B. Time Series in R 2. Time Series Analysis in R

Multivariate Time Series

Multivariate Time Series is creating multiple time series in a single chart.

	positiveCases	deaths
2020.057	580	17
2020.077	7813	270
2020.096	28266	565
2020.115	59287	1261
2020.134	75700	2126
2020.153	87820	2800
2020.172	95314	3285
2020.192	126214	4628
2020.211	218843	8951
2020.230	471497	21283
2020.249	936851	47210
2020.268	1508725	88480
2020.287	2072113	138475

Example:

Taking data of total positive cases and total deaths from COVID-19 weekly from 22 January 2020 to 15 April 2020 in data vector.

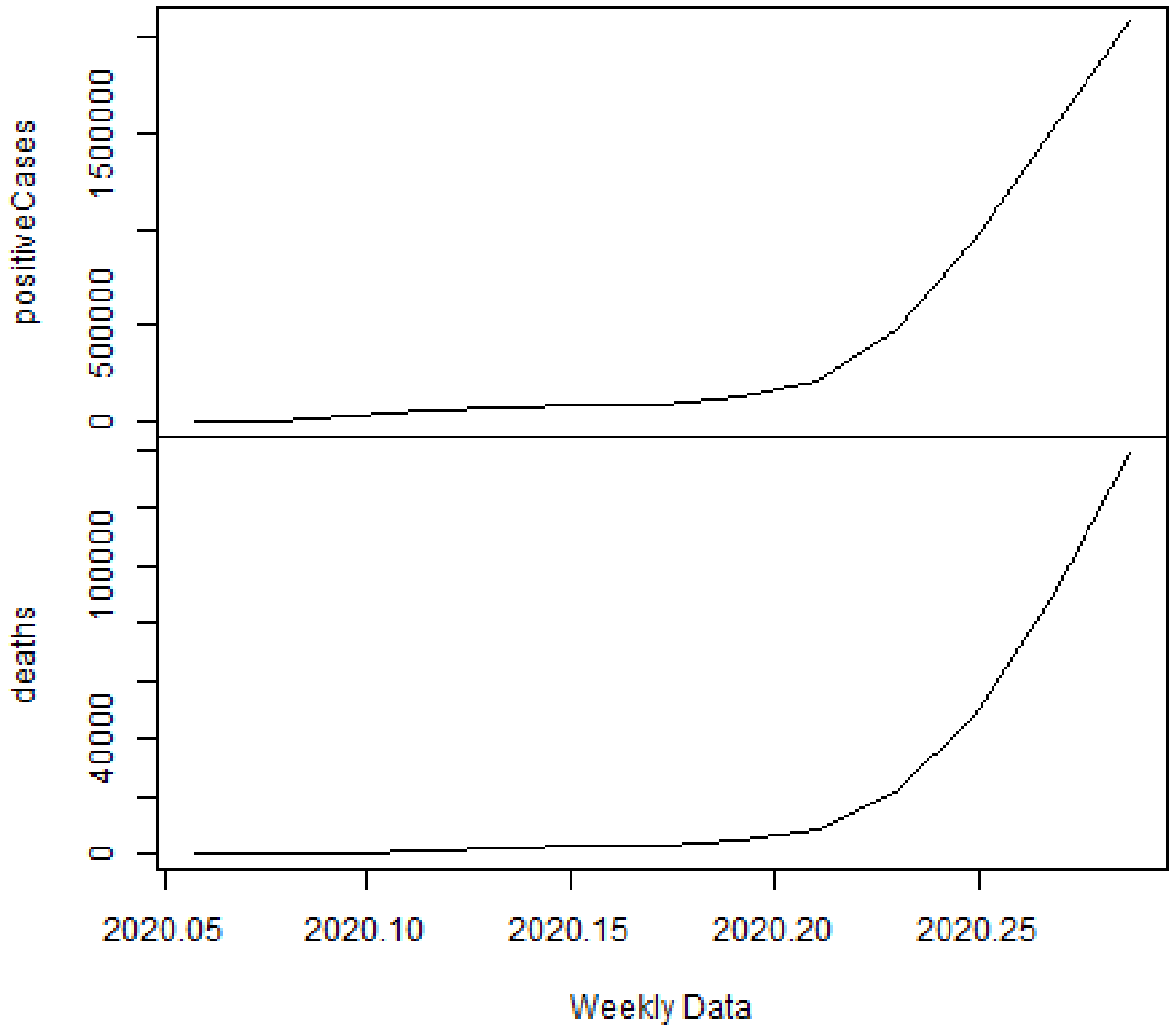
```
positiveCases <- c(580, 7813, 28266, 59287,
                   75700, 87820, 95314, 126214,
                   218843, 471497, 936851,
                   1508725, 2072113)

deaths <- c(17, 270, 565, 1261, 2126, 2800,
            3285, 4628, 8951, 21283, 47210,
            88480, 138475)

# library required for decimal_date() function
library(lubridate)

# output to be created as png file
png(file ="multivariateTimeSeries.png")

# creating multivariate time series object
# from date 22 January, 2020
mts <- ts(cbind(positiveCases, deaths),
          start = decimal_date(ymd("2020-01-22")),
          frequency = 365.25 / 7)
```

B. Time Series in R

2. Time Series Analysis in R

Multivariate Time Series

Multivariate Time Series is creating multiple time series in a single chart.



C. How to visualize time series data

A dashboard for visualizing time series data

- **Running repetitive analytical** queries becomes a force multiplier for organizations looking to expose their time series data across teams.
- **Dashboards** are a great way to visualize and **present time series data** to its target audience in a format that is **meaningful and easy to understand.**

C. How to visualize time series data

Tools for graphing time series data

1. InfluxDB
2. Grafana: https://grafana.com/
3. IoT specific dashboarding tools like Seeq

C. How to visualize time series data

Tools for graphing time series data

1. InfluxDB UI visualization layer

- **Time series data visualization types**

Time series line graphs and bar graphs	Gauge	Heatmap
Single Stat	Graph + Single Stat visualization for time series data	Histogram
Scatter	Gauge	Table

https://www.youtube.com/watch?v=Vq4cDIdz_M8

C. How to visualize time series data

Tools for graphing time series data

2. Grafana

Grafana dashboard examples

Customer overview dashboard

https://play.grafana.org/d/000000110/business-metrics?orgId=1

C. How to visualize time series data

Tools for graphing time series data

2. Grafana

Grafana dashboard examples

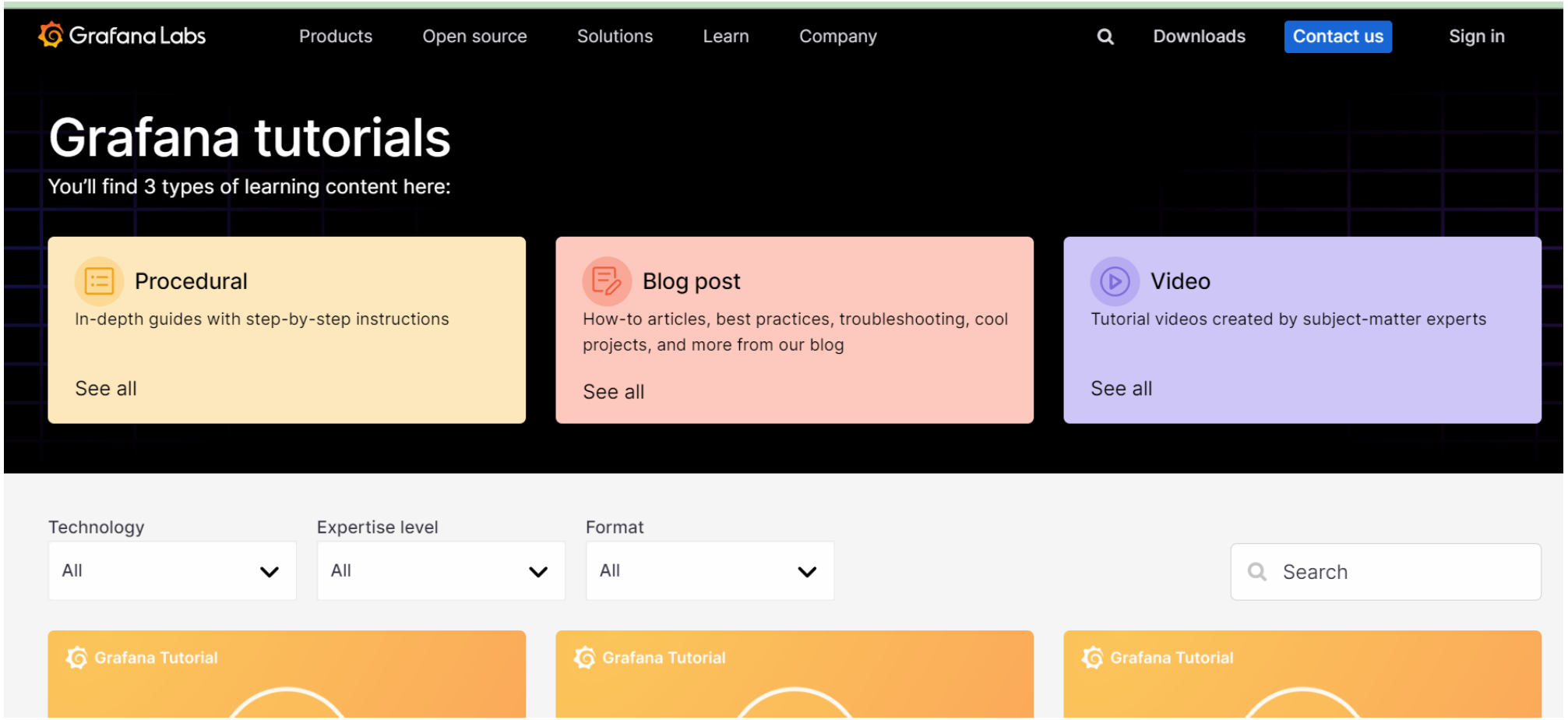
Temperature dashboard

https://play.grafana.org/d/_gFwRQuZks/google-sheets-average-temperature?orgId=1

C. How to visualize time series data

Tools for graphing time series data

2. Grafana



https://grafana.com/tutorials/

C. How to visualize time series data

Tools for graphing time series data

3. Time series custom graphs

Building custom graphs using Dygraphs Charting Library

#install.packages("dygraphs")

- The dygraphs package is an R interface to the dygraphs JavaScript charting library.
- It provides rich facilities for charting time-series data in R, including:
    1. **Automatically plots** xts **time series objects** (or any object convertible to xts).
    2. Highly configurable axis and series display (including optional second Y-axis).
    3. **Rich interactive features** including zoom/pan and series/point highlighting.
    4. Display upper/lower bars (e.g. prediction intervals) around series.
    5. **Various graph overlays** including shaded regions, event lines, and point annotations.
    6. Use at the R console just like conventional R plots (via RStudio Viewer).
    7. Seamless **embedding** within **R Markdown documents** and **Shiny web applications**.

https://rstudio.github.io/dygraphs/

D. Forecasting

Article Examples

1.

Time-series Modeling for **Consumer Price Index Forecasting** using Comparison Analysis of **AutoRegressive** Integrated **Moving Average** and **Artificial Neural Network**

https://www.scitepress.org/Papers/2020/103692/103692.pdf

2.

Forecasting Tehran Stock Exchange index using **the industry index** and **economic variables** affecting it using **neural network**

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.876.2206&rep=rep1&type=pdf

D. Forecasting

Time series forecasting methods

- Time series forecasting uses information regarding historical values and associated patterns to predict future activity.

- Time series forecasting methods include:

1. Trend analysis
2. Cyclical fluctuation analysis
3. Seasonal pattern analysis

D. Forecasting

Time series forecasting methods

```
# library required for decimal_date() function
library(lubridate)

# library required for forecasting
#install.packages ("forecast")
library(forecast)

# output to be created as png file
png(file ="forecastTimeSeries.png")

# creating time series object
# from date 22 January, 2020
mts <- ts(x, start = decimal_date(ymd("2020-01-22")),
          frequency = 365.25 / 7)

# forecasting model using arima model
fit <- auto.arima(mts)
```

```
# Next 5 forecasted values
forecast(fit, 5)

# plotting the graph with next
# 5 weekly forecasted values
plot(forecast(fit, 5), xlab ="Weekly Data",
     ylab ="Total Positive Cases",
     main ="COVID-19 Pandemic", col.main ="darkgreen")

# saving the file
dev.off()
```

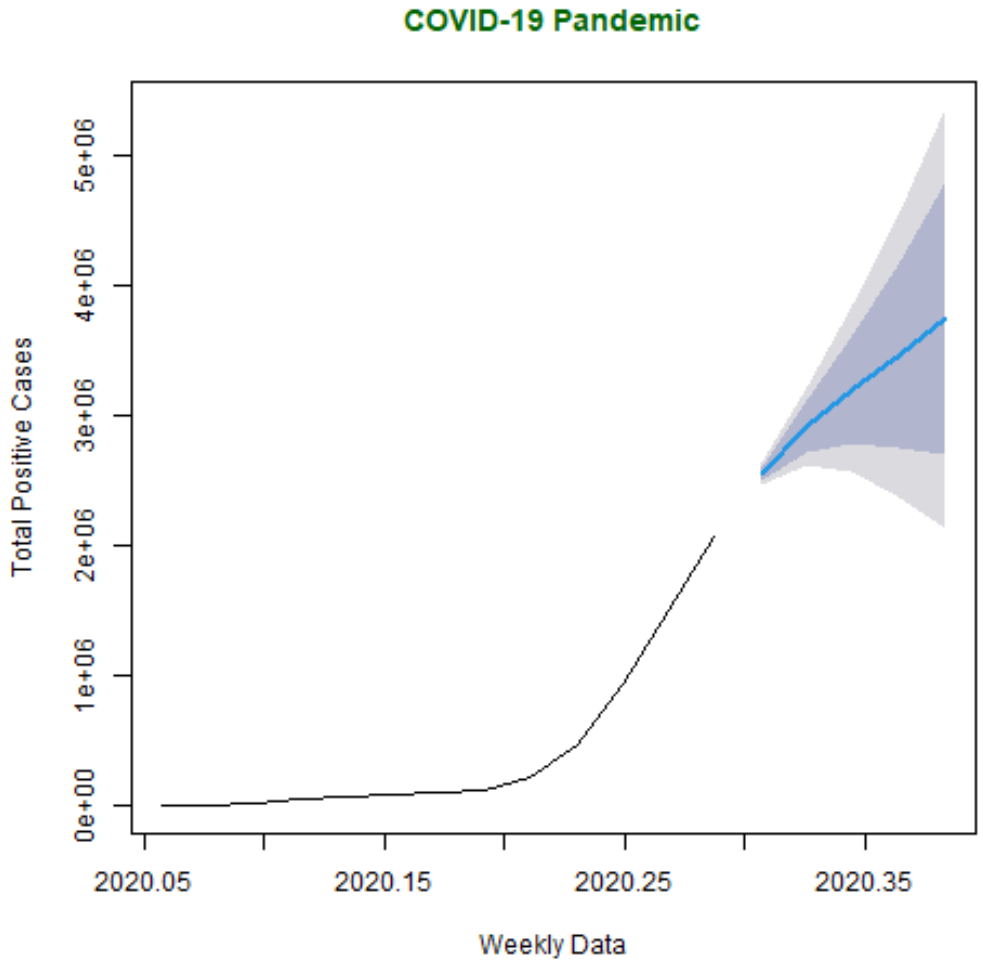
D. Forecasting

Time series forecasting methods

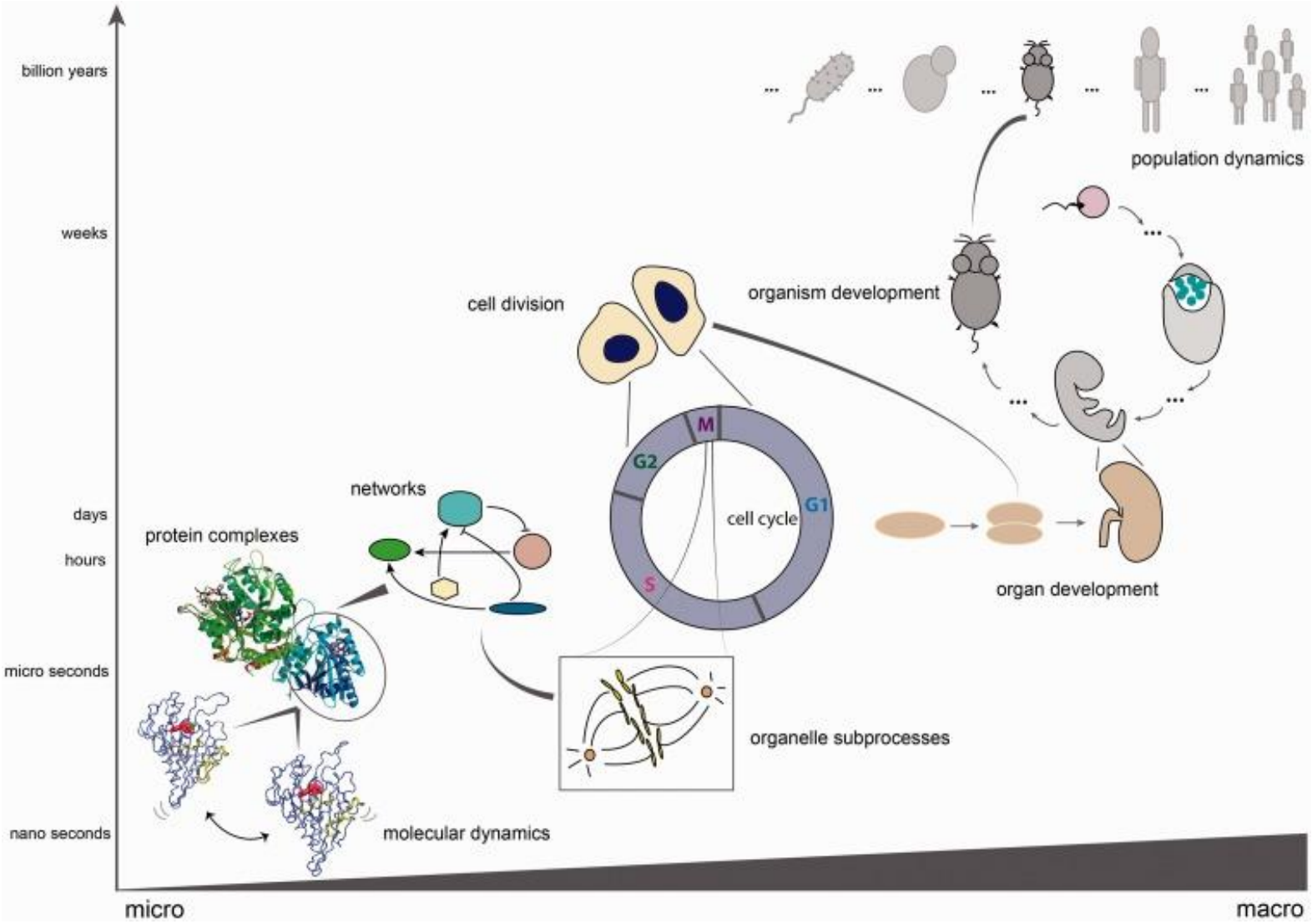
After executing the above code, the following forecasted results are produced.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020.307	2547989	2491957	2604020	2462296	2633682
2020.326	2915130	2721277	3108983	2618657	3211603
2020.345	3202354	2783402	3621307	2561622	3843087
2020.364	3462692	2748533	4176851	2370480	4554904
2020.383	3745054	2692884	4797225	2135898	5354210

graph plots estimated forecasted values of COVID-19 if it continues to be widespread for the next 5 weeks



E. Time series in biology



Different biological time scales

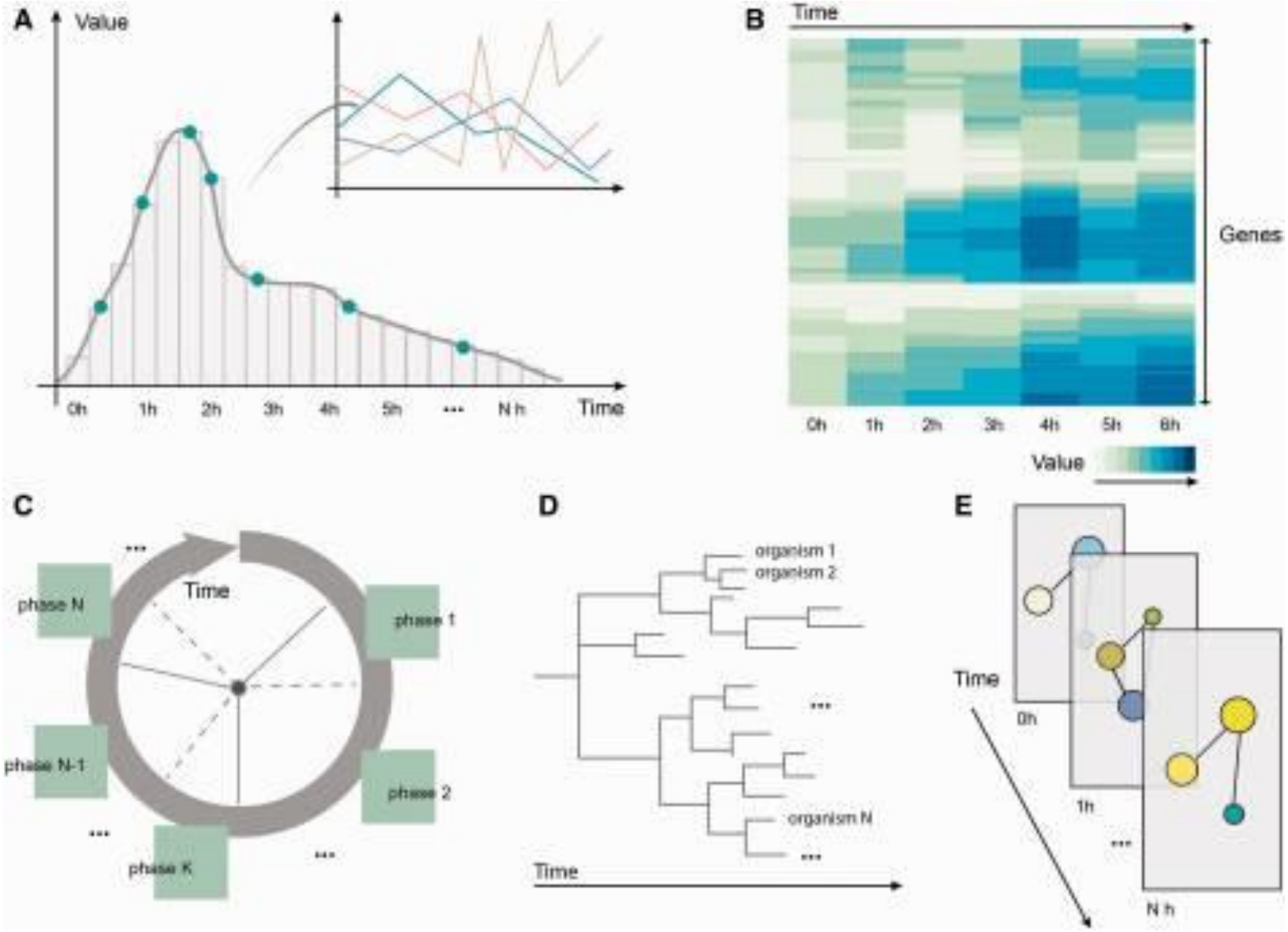
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4171679/figure/bbt021-F1/

E. Time series in biology

- Time is of the essence in biology as in so much else

1. Monitoring disease progression
2. The timing of developmental defects
3. The processes of drug discovery
4. Circadian rhythms

E. Time series in biology



(A.) linear representations

(B) heat maps

(C) circular design

(D) tree-like diagrams

(E) layers

E. Time series in biology

- In the field of biology, one can identify five main approaches to represent time

  (i) linear representations:

  (ii) heat maps

  (iii) circular design

  (iv) tree-like diagrams

  (v) layers

- More complex representations, like splines, contour plots, phase space trajectories, or bifurcation diagrams, build on top of these, many of them inspired by approaches in other fields

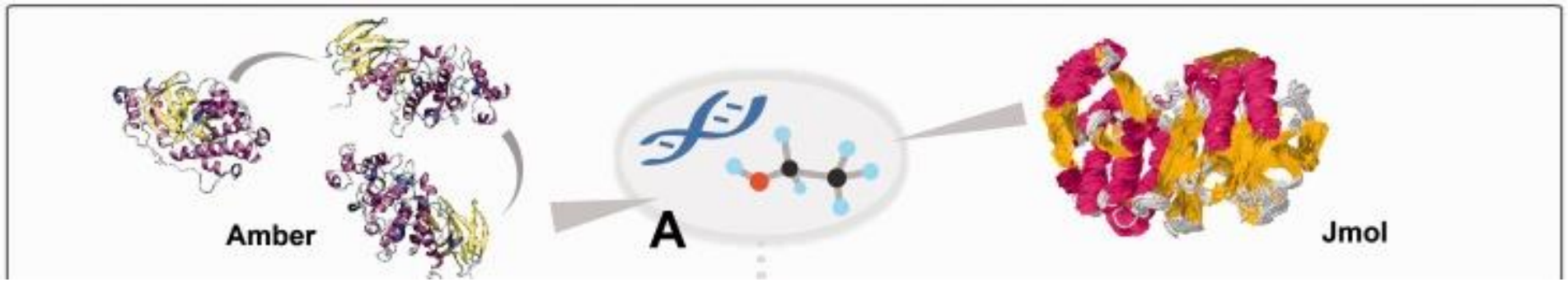
E. Time series in biology

TOOLS FOR REPRESENTING TIME IN BIOLOGY

1. Singular value decomposition (SVD)
2. Principal component analysis (PCA)
3. Self-organizing maps (SOMs)
4. Recurrence quantification analysis (RQA)
5. Fast Fourier transform (FFT)
6. Wavelet decomposition
7. Time warping algorithms

E. Time series in biology

Time at the molecular level



1. Amber
2. Gromacs
3. CHARMM
4. NAMD
5. Desmond: The trajectories of the molecules can then be visualized using tools like

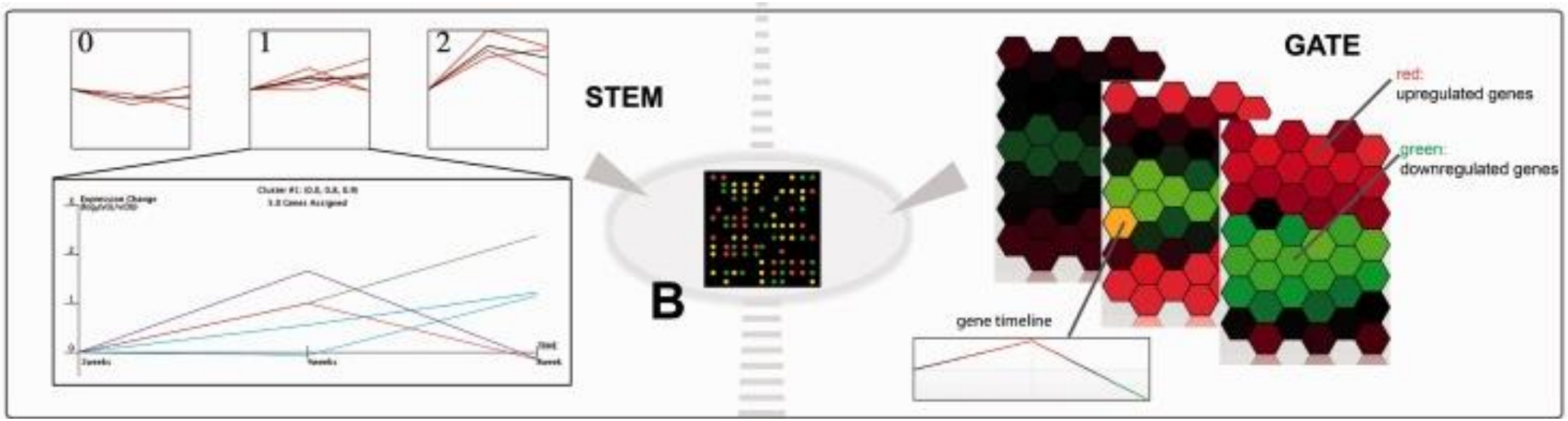
E. Time series in biology

Time at the gene level

1. Heat maps
2. Tools like STEM or XMAS

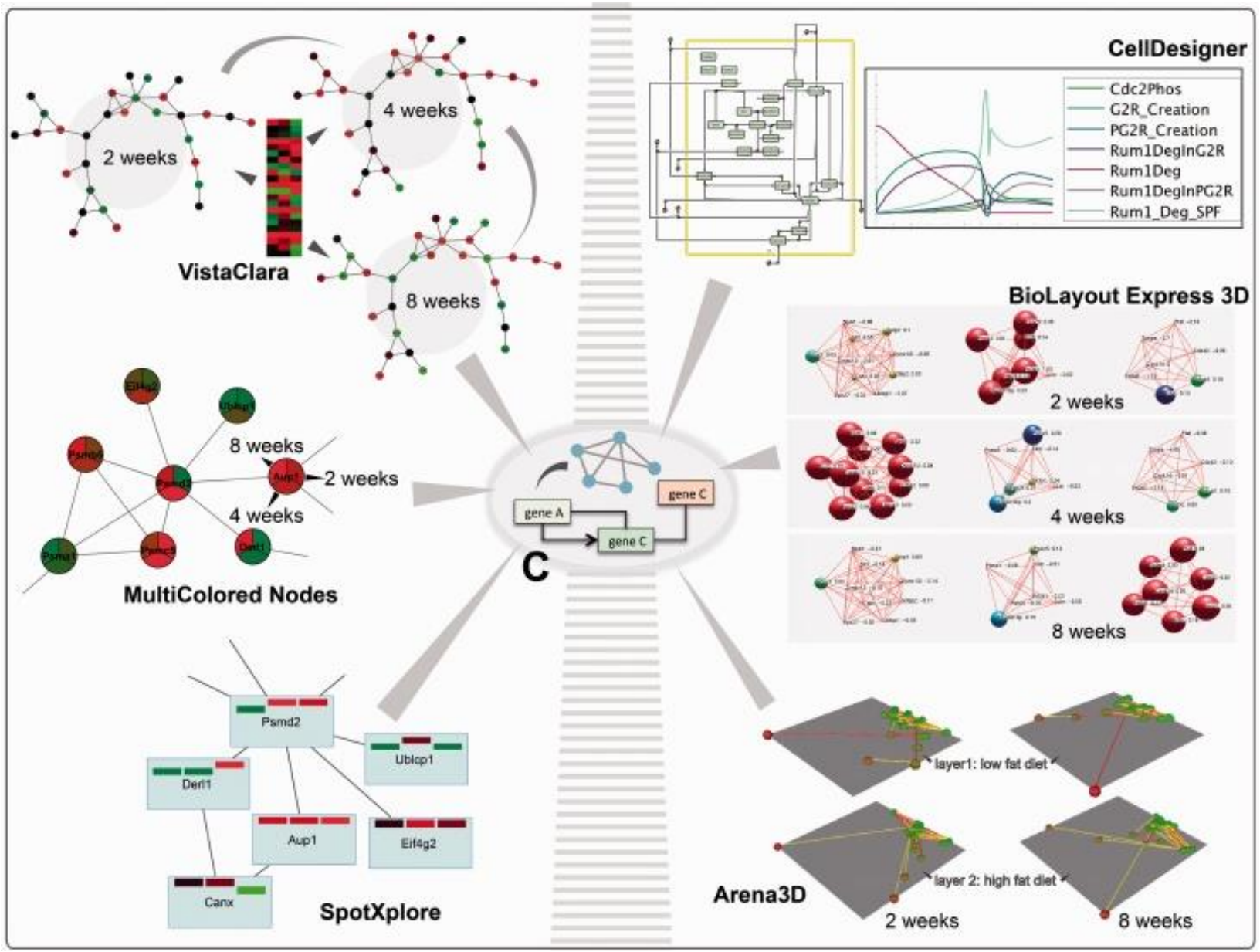
→ provide linear methods to visualize changes and correlations in gene expression patterns, through profile reordering, functional enrichment analysis or multiple trajectory tracking

3. GATE → mimics the microarray setup in a grid of hexagonal cells positioned to denote similarities in gene expression time-course profiles.



E. Time series in biology

Time at the network level



- At the network level, time course changes can be tracked using different Cytoscape plugins, e.g. by animating colour changes in the network with VistaClara, drawing pie chart slices with MultiColored Nodes, or using bar charts embedded in the network nodes with SpotXplore.

E. Time series in biology

At the organismal level

- Shows such depictions for aurora kinase B orthologs in four species.
- In the case of iTOL, additional time course data can be visualized in the form of discs, heat maps or animations (here we show the phases in the cell cycle where this gene has a periodic peak of transcription, as obtained from Cyclebase.

At the organismal level, multiple sequence alignment visualizers, like Jalview, and phylogenetic tree builders, like iTOL, depict evolutionary distances between entities of different organisms.

Take home message

- What is time series data ?
- Why time series visualization is important ?
- What is the best strategy to visualize time series
- Time series ịn R
- Time series in biology

Reference

1. Fry, B. (2008). Visualizing data: Exploring and explaining data with the processing environment. " O'Reilly Media, Inc.".

2. https://campus.datacamp.com/courses/time-series-analysis-in-r/
3. https://www.influxdata.com/what-is-time-series-data/
4. https://www.neonscience.org/resources/learninghub/tutorials/
5. https://www.geeksforgeeks.org/