Visualization for Data Science

Visualization Grammar
Introduction to Altair

Data Abstraction Clicker Review Questions

What is the data attribute type for each column?

postal_code

education_level

IQ_score

pH_preference

birth_year

A. Quantitative – Interval

B. Quantitative – Ratio

C. Quantitative – Temporal

D. Nominal

E. Ordinal

Dataset Description

A psychology research study collected data from 150 participants on cognitive performance and demographics:

Variable	Example Values	Description
postal_code	90210, 10001, 60601, 30309, 94102	Participant's ZIP/postal code
education_level	"High School", "Some College", "Bachelor's", "Master's", "PhD"	Highest education completed
IQ_score	85, 100, 115, 130, 95, 110	Standardized IQ test score
pH_preference	6.2, 7.0, 7.8, 6.5, 8.1, 7.4	Preferred pH level for drinking water
birth_year	1985, 1992, 1978, 2001, 1995, 1988	Year participant was born

Sample Data

postal_code	education_level	IQ_score	pH_preference	birth_year
90210	Bachelor's	110	7.2	1995
60601	Master's	125	6.8	1988
10001	High School	95	7.5	2001
30309	PhD	135	6.5	1982
94102	Some College	105	7.8	1992

Definitions: Marks and channels

- Marks: what we see, the basic geometric primitives
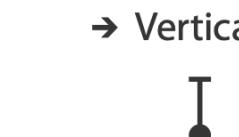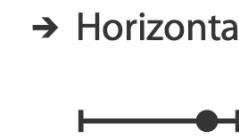
- Channels: the way to control the appearance of marks, independent of the dimensionality of the geometric primitive

- Channel properties differ: type & amount of information that can be conveyed to human perceptual system

→ Points

→ Lines

›› Areas

→ Position
  → Horizontal
  → Vertical
 → Both

→ Color

→ Shape

→ Tilt

→ Size
  → Length
  → Area
  → Volume

3
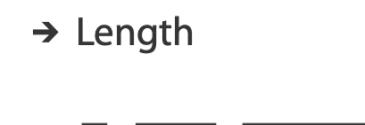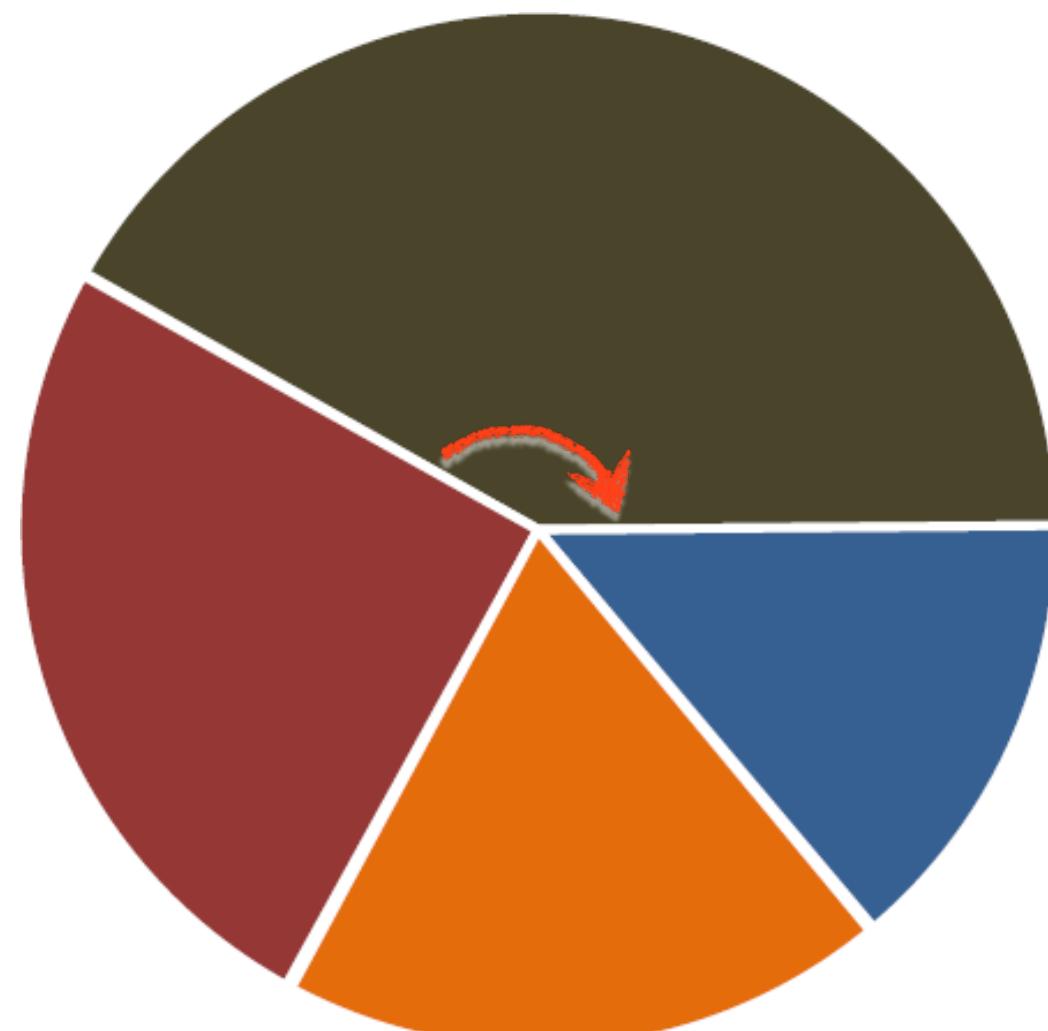
# of attributes encoded: 2

MARK:

- ➔ Points
- ➔ Lines
- ➔ Areas

CHANNEL:

- ➔ Position
  - → Horizontal
  - → Vertical
  - → Both
- ➔ Color
- ➔ Shape
- ➔ Tilt
- ➔ Size
  - → Length
  - → Area
  - → Volume

# Clicker Question

Shooting Media Coverage

What is the mark?

    A: points

    B: lines

    C: interlocking areas

What are the channels?

    A: vertical position

    B: color

    C: horizontal position

    D: area

    E: angle

https://twitter.com/MonaChalabi/status/1158779046693679106?s=20
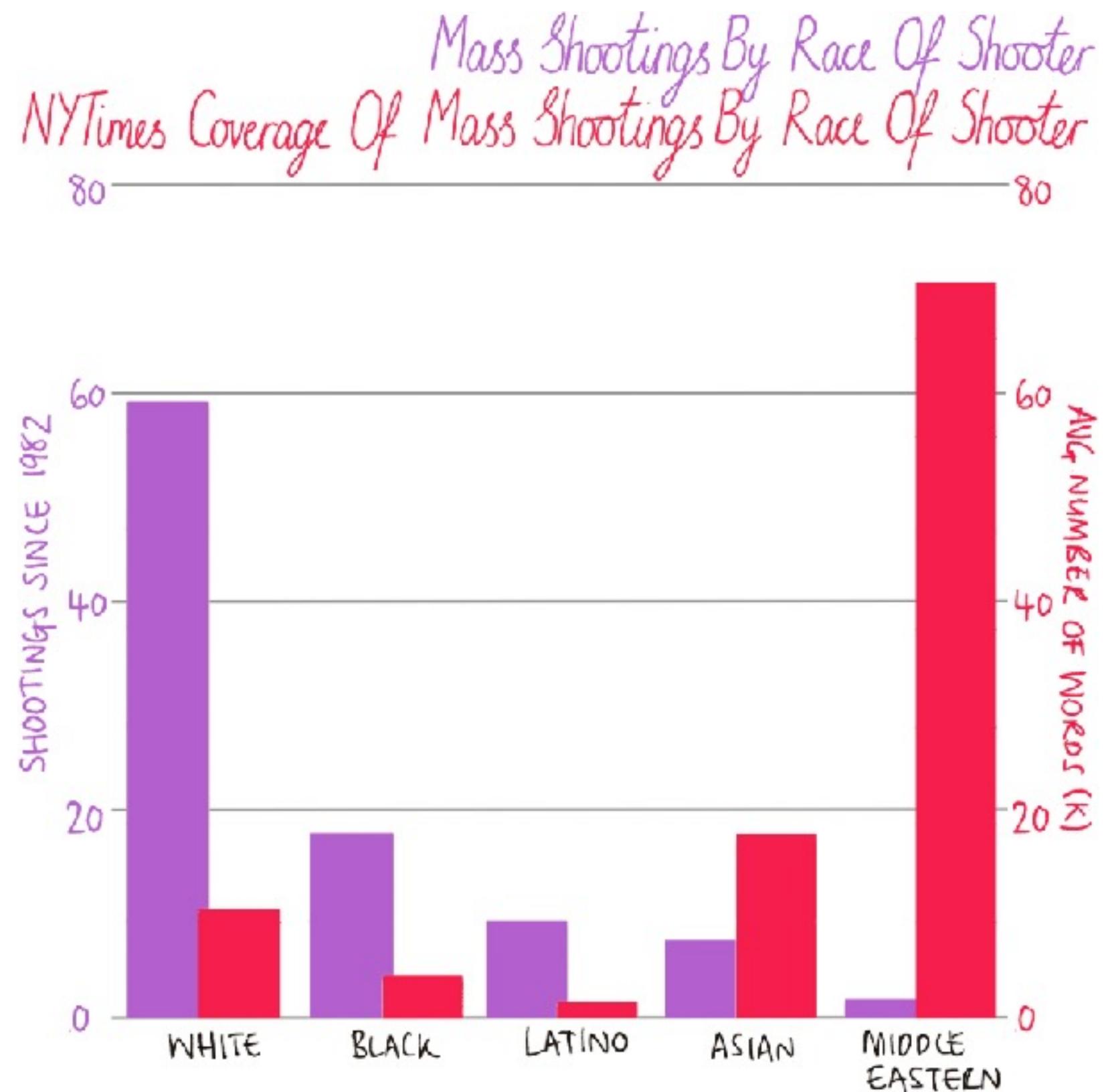
Clicker Question

Alpen Forest Fires
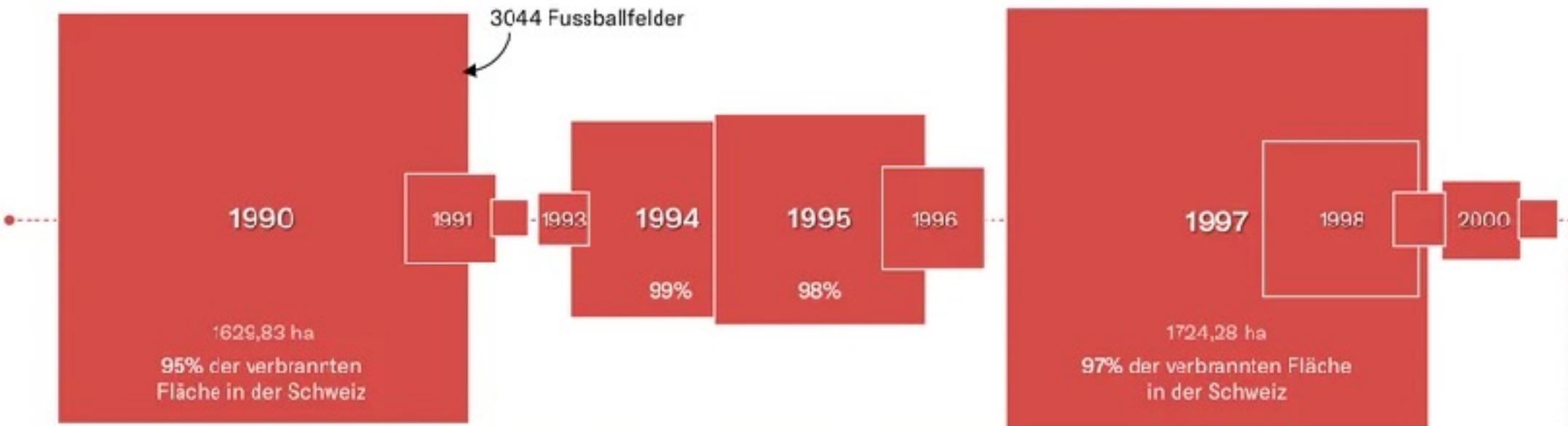
What are the marks

    A: points

    B: lines

    C: interlocking areas

What are the channel(s)

    A: vertical position

    B: color

    C: horizontal position

    D: area

    E: angle

Burned area in hectares on the southern side of the Alps

3044 Fussballfelder

1990 1991 1993 1994 1995 1996 1997 1998 2000

99%   98%

1629,83 ha
95% der verbrannten
Fläche in der Schweiz

1724,28 ha
97% der verbrannten Fläche
in der Schweiz

Clicker Question

More Alpen Forest Fires

What are the mark(s)
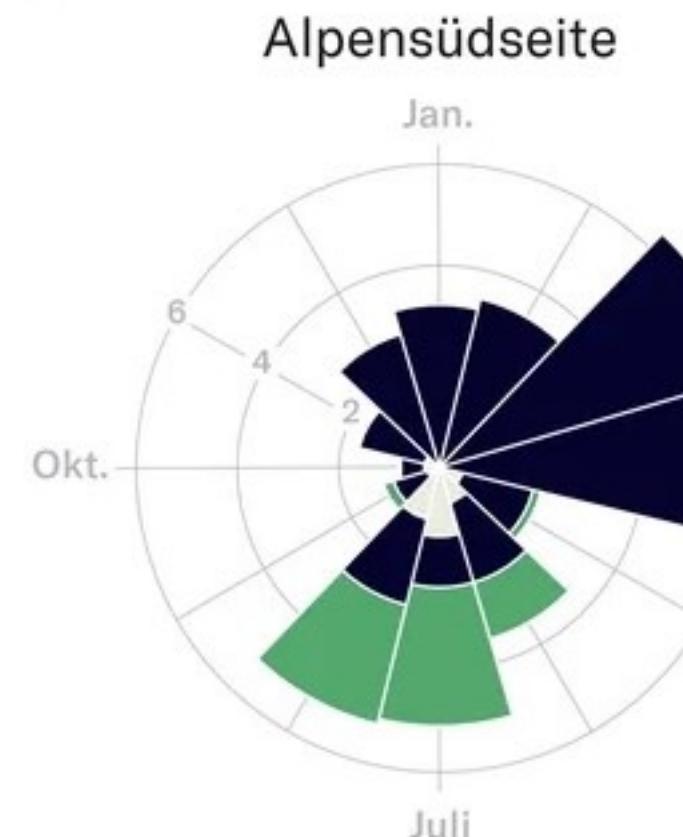
    A: points

    B: lines

    C: interlocking areas

What are the channel(s)
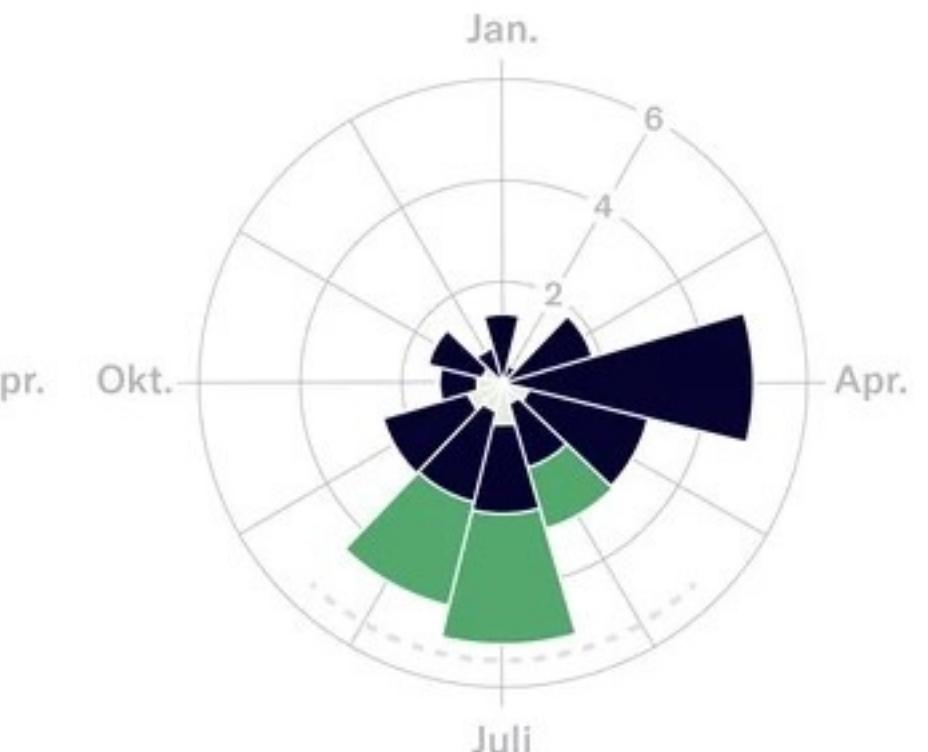
    A: position

    B: color

    C: length

    D: area

    E: angle

Monthly distribution of forest fires in the Alpine regions caused by. , ,

● den Menschen   ● Blitzschläge   ○ unbekannt
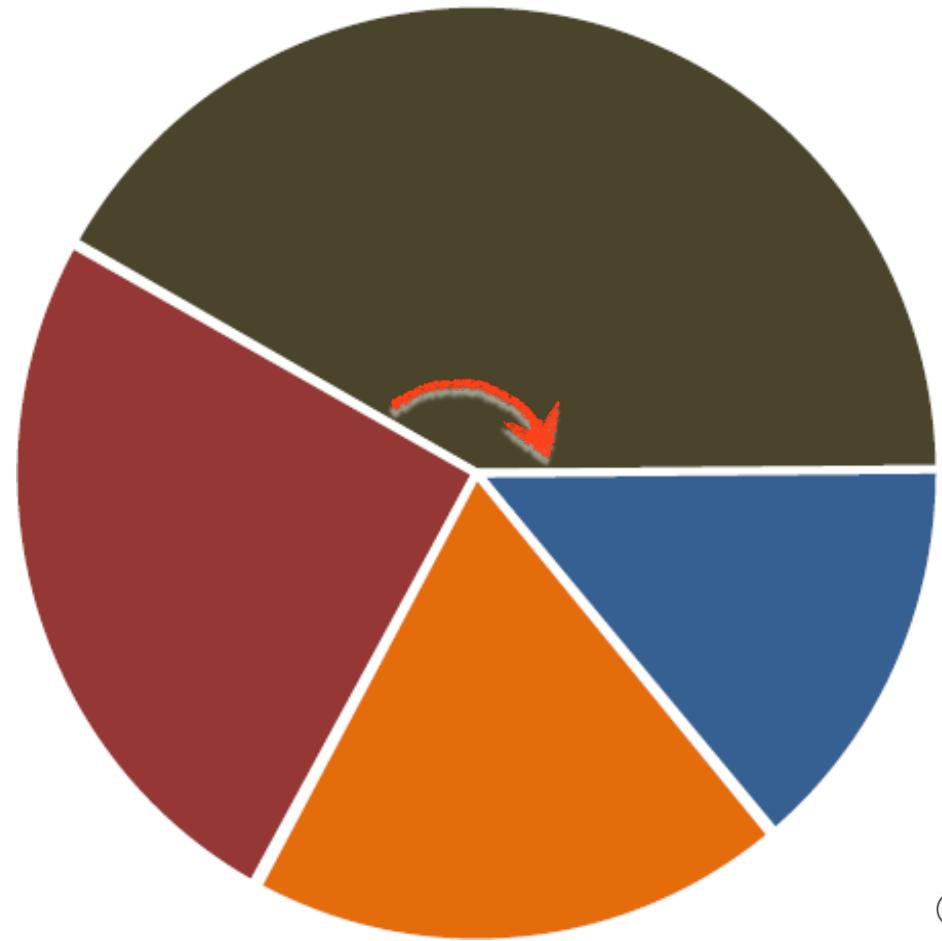
Alpensüdseite   andere Alpengebiete

Average numbers in the period 2000-2018
Source: Swissfire forest fire database

NZZ / awi.

Position
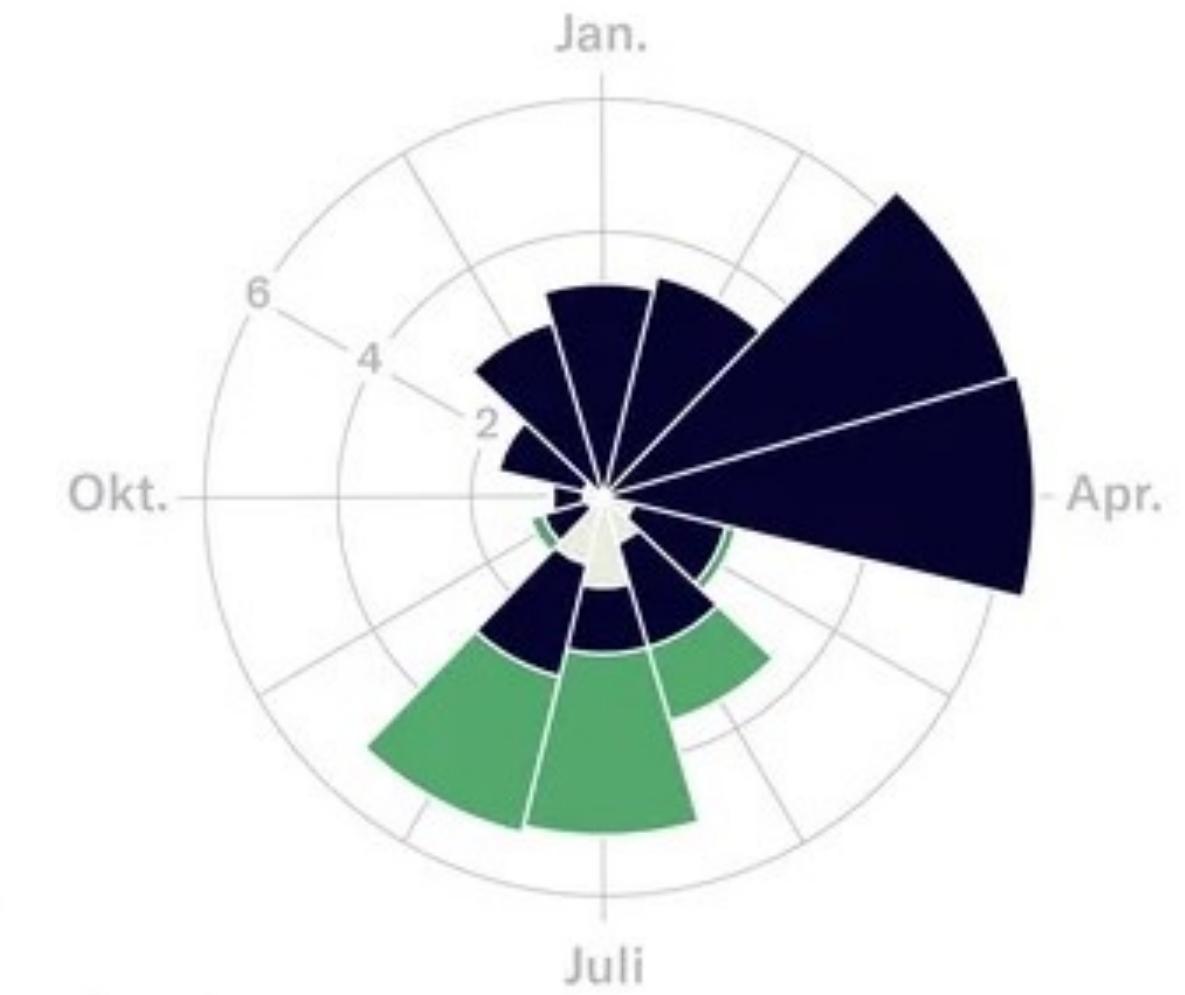
→ Horizontal → Vertical → Both

Color

Shape

Tilt

Size

→ Length   → Area   → Volume

Learning Outcomes

Use makes and channels to create charts in Altair

- Chart Object

- How to attach data to the Chart

- How to specify each channel and describe which attribute encodes
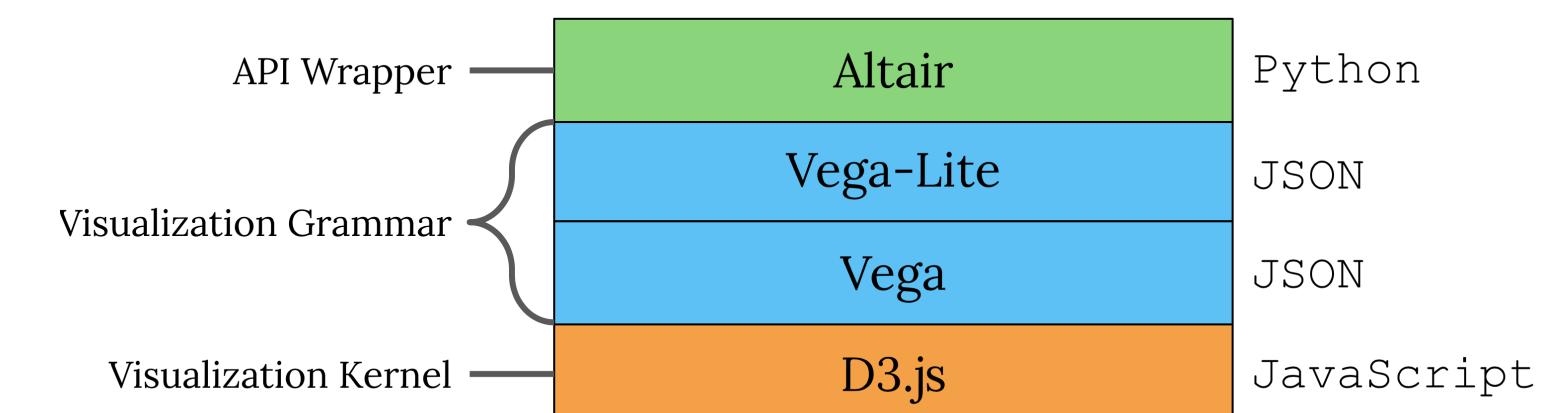
Grammar for Creating Visualizations

A framework which follows a layered approach to describe, design, create visualizations in a structured fashion.

Grammar of Graphics – Leland Wiklinson Layered Grammar of Graphics – Hadley Wickman (R visualization package ggplot2)
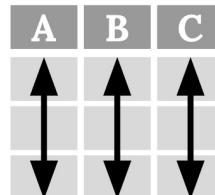
Vega is a visualization grammar, a declarative language for creating, saving, and sharing interactive visualization designs. Vega-Altair – statistical visualization library for Python.

Others: Cicero, Mascot

API Wrapper	Altair	Python
Visualization Grammar	Vega-Lite	JSON
	Vega	JSON
Visualization Kernel	D3.js	JavaScript

What is Altair

https://www.youtube.com/watch?v=U7w1XumKK60
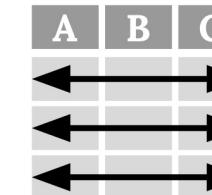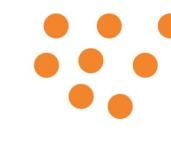
Variables & **Observations**

Line

Circle
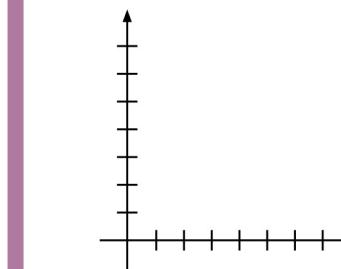
A A B / A C B B / A C C B / C C
Text
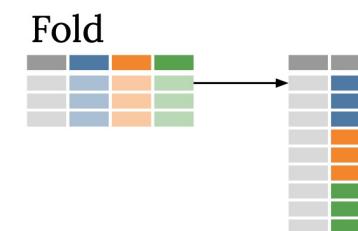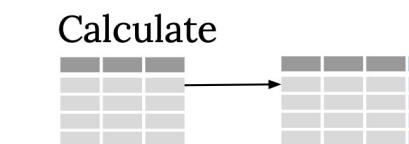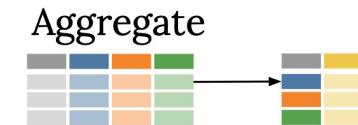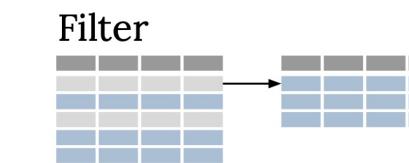
Bar

Channel	Variable
X Position →	A
Y Position →	B
Size	C
Color	D
⋮	⋮

Guide
Scale
Transform
Encoding
Mark
Data

Legend
● A
● B
● C

$$f(\texttt{domain}) \longrightarrow \texttt{range}$$

Filter

Aggregate

Calculate

Fold

Altair Basics

Create a Chart Object

Attach data to the Chart Object

Specify the mark type

```
alt.Chart(data).mark_bar().encode(
        channel_1 = 'column1',
        channel_2 = 'column2',
)
```

Specify each channel and what data attribute it encodes