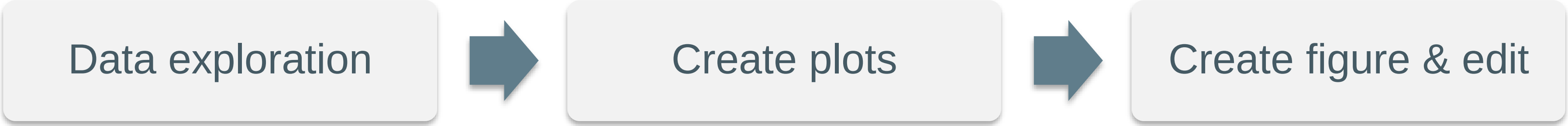
Data Science
Survival Skills

Exercise 4 – Data exploration and visualization

Today‘s workflow



Data exploration

Create plots

Create figure & edit

pandas

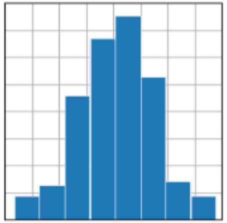
matplotlib

+

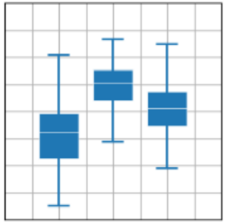
seaborn

INKSCAPE

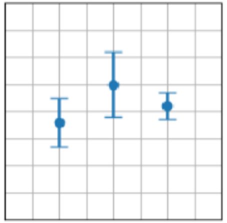
Plotting distributions



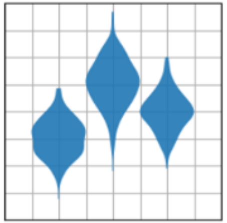
hist(x)



boxplot(X)



errorbar(x, y, yerr, xerr)



violinplot(D)

matplotlib

histplot(df, x)

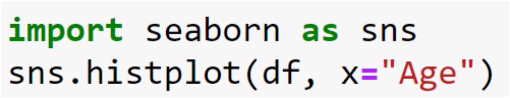
boxplot(df, x)

pointplot(df, x)

violinplot(df, x)

seaborn

DataFrame



```
import seaborn as sns
sns.histplot(df, x="Age")
```

Visualization

understand the **context**

choose an **effective visual**

eliminate **clutter**

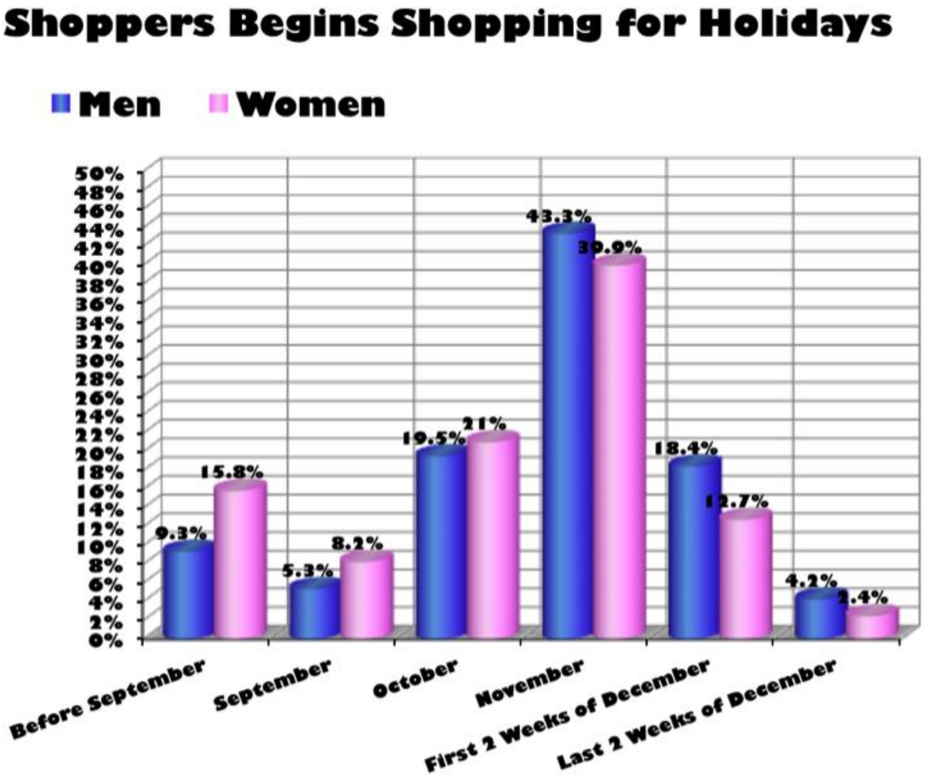
focus **attention**

DATA?
SET UP
ACT!
TENSION
TIME

tell a **story**

Image source: https://www.storytellingwithdata.com/

Removing clutter and focusing attention


Shoppers Begins Shopping for Holidays

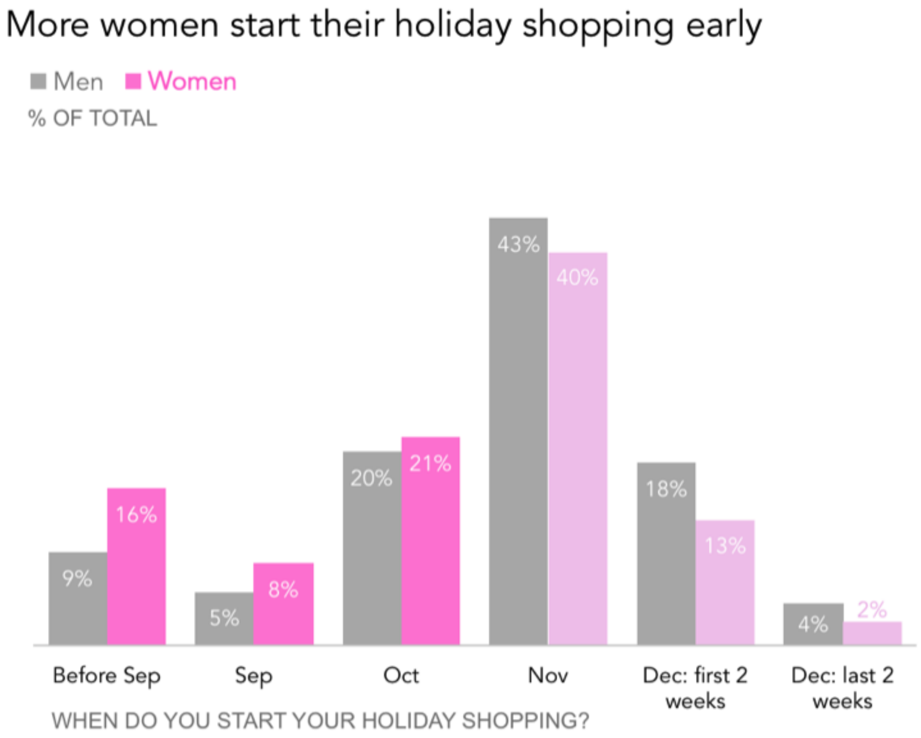
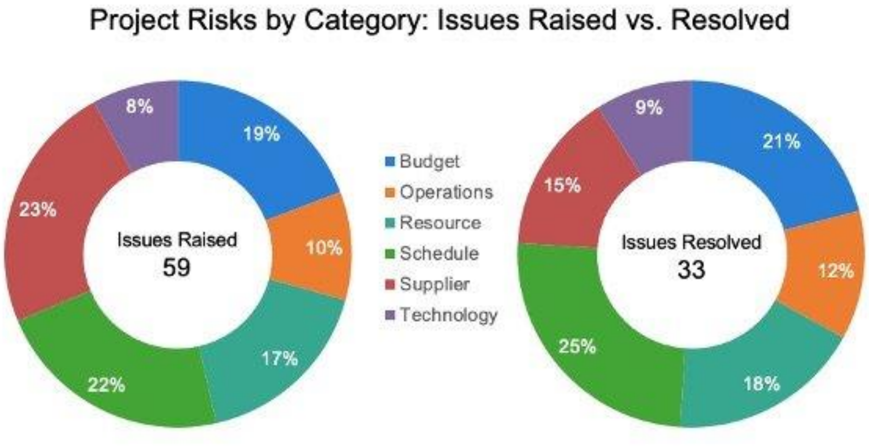

More women start their holiday shopping early

Image source: https://www.storytellingwithdata.com/

Removing clutter and focusing attention

BEFORE



AFTER

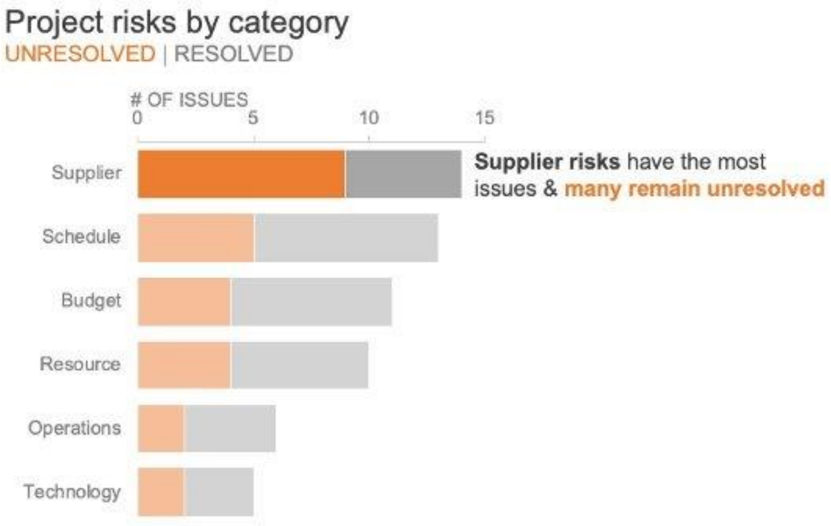


Image source: https://www.storytellingwithdata.com/

Some coding

Raster graphics vs. vector graphics

	Raster	Vector
File extensions	**.png** .jpg .gif	**.svg** .ai .emf
Built from …	Pixels	Mathematical equations, lines, and curves
Usage	Photos, presentations, web, …	Figures for scientific papers, illustrations, logos, …
Pros and cons	- Lose quality when resized + compatibility	+ Don‘t lose quality when resized (very scalable)

Example: Scientific Figure

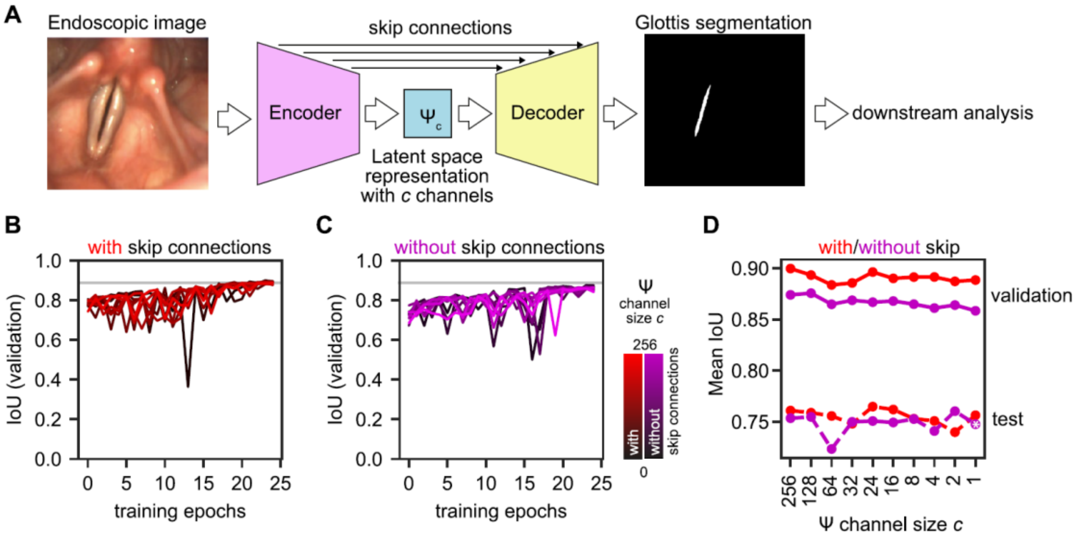


Figure 1. A single latent space channel is sufficient for glottis segmentation. (**A**) Glottis segmentation of endoscopic images using deep neural networks (DNNs) with latent space Ψ. (**B**) Convergence of segmentation DNNs across different latent space channels with enabled skip connections. Gradient from black to red indicates fewer channels. The gray line indicates maximum IoU score. (**C**) Convergence of segmentation DNNs across different latent space channels with disabled skip connections. Gradient from black to magenta indicates fewer channels. The gray line indicates maximum IoU score from panel **B**. (**D**) Performance of best performing segmentation DNNs on validation set (solid lines) and evaluated on test set (dashed lines) with enabled (with, red) and disabled (without, magenta) skip connections across latent space (Ψ) channels measured by mean intersection over union (IoU). The asterisk indicates the architecture used in the subsequent experiments.

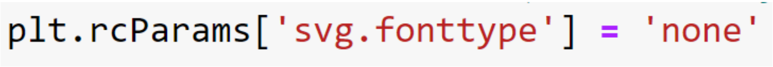
Reference: Kist, Andreas M., et al. "A single latent channel is sufficient for biomedical glottis segmentation." *Scientific Reports* 12.1 (2022): 14292.

Editing plots in Inkscape

Download: https://inkscape.org/

Tutorials: https://inkscape-manuals.readthedocs.io/

Make font editable:



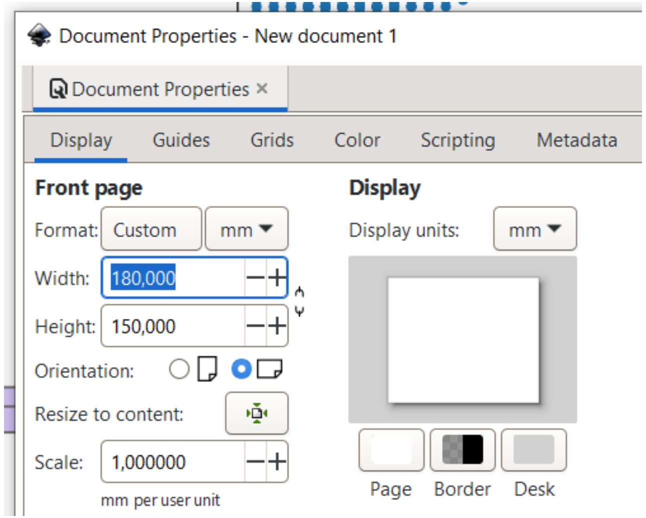
```
plt.rcParams['svg.fonttype'] = 'none'
```

Inkscape

INKSCAPE

Set figure size:

File → Document properties



Document Properties - New document 1

Document Properties ×

Display | Guides | Grids | Color | Scripting | Metadata

Front page

Format: Custom mm

Width: 180,000

Height: 150,000

Orientation: ○ ●

Resize to content:

Scale: 1,000000

mm per user unit

Display

Display units: mm

Page Border Desk

Snap to grid points

View → ☑ Page Grid

View → Show/Hide → Snap controls bar → Activate snapping

Aligning elements

Shift + Ctrl + A

Inkscape

Save as SVG (vector graphic)
File → Save as …

Export as PNG (raster graphic)
File → Export …

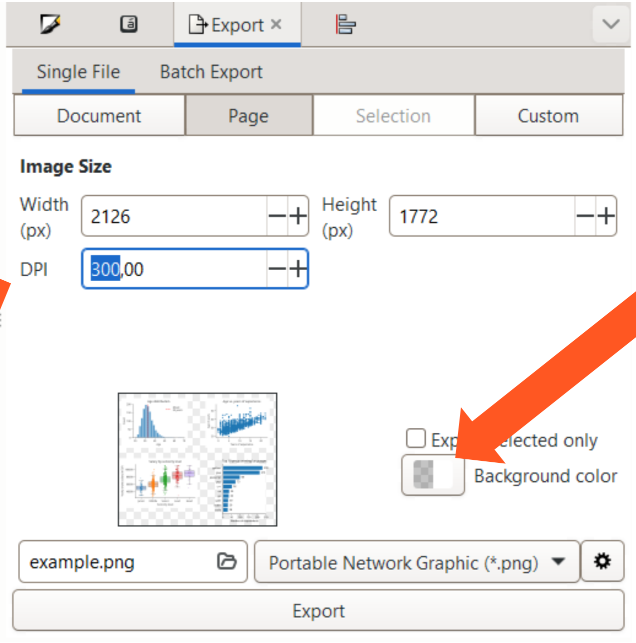


Image Size

Width (px): 2126

Height (px): 1772

DPI: 300,00

□ Exp elected only

Background color

example.png

Portable Network Graphic (*.png)

Resolution
(dots per inch)

Set background
to white or
transparent

More coding

Questions?

And we are done!

Thank you