AIR QUALITY IN RELATION TO PERFORMANCE

DVIZ Project HS22

Gurme Chodon
Kramer Lukas
Nyungmartsang Choekyel

9. Januar 2023
Suurstoffi 1, 6343 Rotkreuz

MOTIVATION

Air pollution is one of the major environmental concerns that exists, and it has made its way into the offices. It has become prevalent, and people are getting more tired, losing focus on their work, as well as being less productive due to high CO_2 concentration in small spaces. With that being said, we aim to raise awareness about the potential consequences that indoor air pollution has on people's productivity.

WORK DESCRIPTION

Our first step was to understand the data we received from the architecture's office of GKS[1]; therefore, we went on-site and observed two office floors and took notes on the equipment: amount and localization of plants, air-Q devices, and count of workspaces. Regarding similar works, there are a lot of analyses on air pollution. Nevertheless, the structure and encodings were all mostly similar. They used line charts, heatmaps and rigid lines for tracking behavior over a given period.[2] We further compared the data from the air-Q devices with studies about indoor pollution and understood the purpose of the measurements.[3] [4] [5]

Later, we cleaned and transformed the raw data of the air-Q and with that, we were ready to start with the visualizations.

[1] (GKS 2023)
[2] (Kuniyoshi 2021)
[3] (Harvard-Studie 2021)
[4] (air-Q 2022)
[5] (air-Q 2022)

USED CHART TYPES

Selecting our colors for data visualization was not an easy task. We had to focus on different color palettes, avoid pure, bright, and highly saturated colors and be consistent with colors across the charts.[6] We also consider readers with colorblindness with the help of the colorblindness simulator.[7] As a result, we narrowed down our color options into the following:

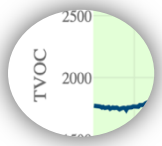
Color	HEX	Description
	#4C90BA	CO_2
	#BDBD00	TVOC
	#F4B811	Temperature
	#E87000	Performance

Color	HEX	Description
	#5D2749	Office 1
	#E57872	Office 2

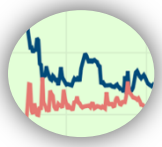
Furthermore, we focused on using the same font "Courier New, monospace" throughout the visualization and applied different font sizes to emphasize different parts. After defining our colors, we standardized our templates, annotations, and layouts, which we utilized throughout the charts.[8]

Let's dissect our chart:

For the **X-axis** we decided to use work hours between 05:00 to 21:00 because most movements occur around this period.



As for the **Y-axis** we decided to take various measurements into consideration. For some we measured in absolute values and in other maximum and minimum values. To display them together.



We mapped both attributes by **position**, which allowed us to learn and interpret new possibilities about the data and guide the reader through our findings.[9] [10]

[6] (Muth 2020)
[7] (Wickline 2006-2021)
[8] (plotly 2022)
[9] (plotly 2022)
[10] (tableau 2022)

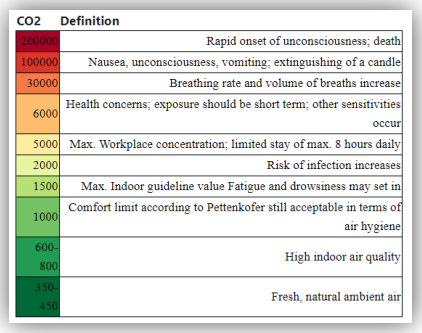
8'447 characters (with spaces)

Since we are looking for trends over time, we decided to use line charts, because it helps to reflect the total values and fluctuations very understandably. Additionally, we used heatmaps for quick impressions for the readers. In some cases, we even combined the heatmap and the line to convey the data in two different ways. Our goal was to create charts that generate value and achieve their purpose in an easily recognizable way.

Most of the time we used **comparison** charts to compare magnitude of values from office 1 and 2. The charts show average values over time and the process gradually.

We also used **distribution** charts to see the measurement values distributed along the time axis. The viewer can identify characteristics such as the range of values, central tendency, shape, and outliers by looking at the shape of distribution.

Unfortunately, there is not a one-size-fits-all chart for timed data set, which is why some situations in our data story dictate line charts while others, where we show a good overview, merit area charts. As for the area chart, we compared the intensity of the average measurements by using color shading.

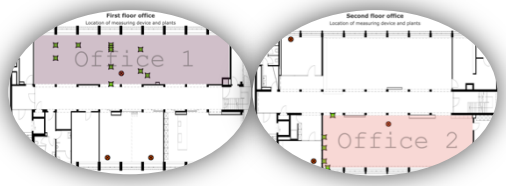


CO2	Definition
200000	Rapid onset of unconsciousness; death
100000	Nausea, unconsciousness, vomiting; extinguishing of a candle
30000	Breathing rate and volume of breaths increase
6000	Health concerns; exposure should be short term; other sensitivities occur
5000	Max. Workplace concentration; limited stay of max. 8 hours daily
2000	Risk of infection increases
1500	Max. Indoor guideline value Fatigue and drowsiness may set in
1000	Comfort limit according to Pettenkofer still acceptable in terms of air hygiene
600-800	High indoor air quality
350-450	Fresh, natural ambient air

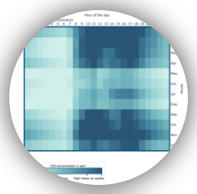
The table provides aid to the readers in understanding the CO_2 values better. The source of this table is from a picture we found on the internet. The picture contained information we thought would be good to know for the reader of our story. To not just repost the picture, we thought of a way to visualize the information in an easy-to-understand table.[11]

[11] (CIK-Solutions 2022)

8‘447 characters (with spaces)



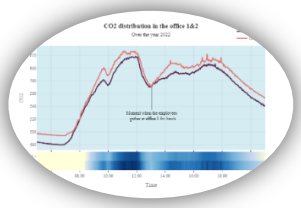
The idea behind both floor plans is to provide an overview of the location of measuring device and plants for the readers. Furthermore, both the colors “purple” and “pink” helps distinguish the offices.

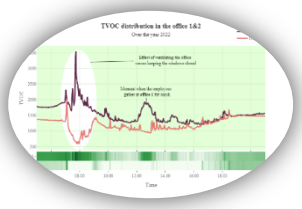


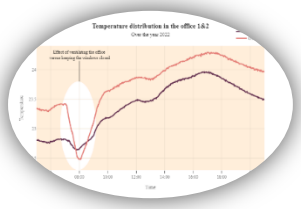
We used a heatmap to show the aggregated overview into a custom area. When areas overlap, we defined colors to indicate the density.

We created this plot to give the reader an idea about our topic and to show how it could look in a worst-case scenario inside a meeting room.

Because of the clearly visible correlations of the three plots, it is quite easy to explain the dependencies of each subplot. For example, the marked area shows the opening of the windows, which results to an improvement of the air quality. To give readers an even better impression of how to understand the plots, we inserted dotted-lines with the same information as the CO_2 table in the first subplot.

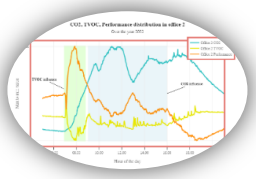






In these charts we tried to show the readers the kind of influence plants and air quality has. Additionally, the measurements of a good air quality have been depicted in the chart. Special events that influence our plots are annotated so that the reader understands the irregularities. Topics that are mentioned in the text are marked with a transparent area. The background color clarifies each attribute’s color, which we defined throughout the plots and the heatmap shows the density of each measurement.

8‘447 characters (with spaces)



Based on this plot, we wanted to explain to our readers, the correlation between TVOC and CO_2 as well as the performance of workers. For an easy explanation, we colored the areas and inserted annotations to show where correlation is clearly visible.

This is the plot where readers can find the correlations themselves. It comes with some presets to make this intention more visible. Because there are many lines, we used the same colors for the attributes as we used in the plots before.

Regarding this plot, we show the readers our result of the discussed topic based on the given example. To make the difference between the two offices more evident, we filled the areas to give a clear impression of the impact. We annotated the most meaningful statement and described the rest in our data story.

TARGET AUDIENCE

This report is for everyone who works indoors. For example, normal office workers and students. We used line charts for our readers because they are familiar with seeing behavior and progression over time. They understand how a line chart works due to their daily work on how to connect information and are aware of outliers.

LIBRARIES AND PACKAGES

Quarto

We used Quarto because it allows Jupyter notebooks to be rendered in a wide variety of formats (eg. Html, pdf) and has great deal of control over code execution, formatting, and appearance.[12]

[12] (quarto 2022)

Plotly

We applied Plotly because it takes little code to produce a visually appealing and interactive chart. The visualizations made through Matplotlib for example are not interactive and looks much worse than default Plotly charts.

NumPy

NumPy is fundamental to Python. We used it for scientific computing and enhancing our speed of execution.

Pandas

Functionality of Pandas is built on top of NumPy, which means that it is suitable to use both together. We used Pandas to analyze data without the need for a database.

Copy

We used deep copy to create a fully independent clone of the original.

DATA SOURCE

As data source we decided to use real time data from our friend, who works in an architecture office for GKS, which extracts the data from their air-Q sensors.

CONCLUSION

Overall, it was an interesting experience for us to work with the data from air-Q and GKS. We have found that windows should be opened regularly to decrease CO_2 and TVOC particles in the room. Under those circumstances the performance increase and health benefit are guaranteed. Additionally, to opening windows, having more plants in an office also has a positive impact on CO_2 concentration.

Therefore, CO_2 concentration should be monitored constantly, and measures should be taken accordingly. Otherwise, the performance of workers will reduce. We hope that our data story will make our readers more aware of this issue.

8‘447 characters (with spaces)

BIBLIOGRAPHY

air-Q. *airQ.* 2022. https://www.air-q.com/produktvarianten.

—. *airQ.* 2022. https://www.air-q.com/gesundheit-und-leistung-der-luftqualitaet.

CIK-Solutions. *cik-solutions.* 2022. https://www.cik-solutions.com/anwendungen/co2-im-innenraum/.

GKS. *GKS.* 2023. https://www.gks.ch.

Harvard-Studie. *mdr wissen.* September 10, 2021. https://www.mdr.de/wissen/schlechte-luft-schlechter-arbeiten-100.html.

Kuniyoshi, Andre. *Analytics Vidhya.* July 20, 2021. https://www.analyticsvidhya.com/blog/2021/07/how-to-analyse-air-pollution-using-python/.

Muth, Lisa Charlotte. *Datawrapper.* September 4, 2020. https://blog.datawrapper.de/beautifulcolors/.

plotly. *plotly.* 2022. https://plotly.com/python/templates/.

—. *plotly.* 2022. https://plotly.com/python/time-series/.

quarto. *quarto.* 2022. https://quarto.org/docs/tools/jupyter-lab.html.

tableau. *tableau.* 2022. https://www.tableau.com/data-insights/reference-library/visual-analytics/charts/line-charts.

Toodu, Eneli. *Analyzing Air Quality Data using Python Pandas and Plotly.* Januar 6, 2021. https://medium.com/planet-os/analyzing-air-quality-data-from-planet-os-datahub-using-python-pandas-and-plotly-f2766c003c6c.

Wickline, Matthew. *Colblindor.* 2006-2021. https://www.color-blindness.com/coblis-color-blindness-simulator/.

8‘447 characters (with spaces)