



**Vilniaus
universitetas**

Informatikos ir informatinio mąstymo mokomoji veikla

Duomenys. Duomenų tyryba



Kuriame
Lietuvos ateitį
2014–2020 metų
Europos Sąjungos
fondų investicijų
veiksmų programa



**Vilniaus
universitetas**

Duomenys. Duomenų tyryba

Informatikos ir informatinio mąstymo mokomosios veiklos sukurtos įgyvendinant projektą „Aukštųjų mokyklų tinklo optimizavimas ir studijų kokybės gerinimas Šiaulių universitetą prijungiant prie Vilniaus universiteto“ (Nr. 09.3.1-ESFA-V-738-03-0001), finansuojamą iš Europos socialinio fondo lėšų pagal 2014–2020 metų Europos Sąjungos fondų investicijų veiksmų programos 9 prioriteto „Visuomenės švietimas ir žmogiškųjų išteklių potencialo didinimas“ įgyvendinimo priemonę Nr. 09.3.1-ESFA-V-738 „Aukštųjų mokyklų tinklo tobulinimas“.

Metodinė medžiaga „Duomenys. Duomenų tyryba“ skirta Mokyklos pedagogikos studijų programos moduliui „Informatikos didaktika“. Tikslinė grupė – būsimi informatikos pagrindinio ugdymo mokytojai. Medžiaga siejama su informatikos ir matematikos Bendrosiomis programomis, duomenų tyrybos ir informacijos informatikos pasiekimų sritimi. Atlikdami veikloje numatytas užduotis studentai, būsimieji mokytojai, išsiaiškins duomenų tyrybos pagrindus ir jų mokymo mokykloje aspektus. Pateikiamas teorinis temos pagrindimas būsimam mokytojui, aptariamos pagrindinės srities sąvokos, siūlomi klausimai diskusijoms, užduotys praktiniam darbui.

Veiklą parengė: dr. Tatjana Jevsikova

Redagavo: Viktoras Dagys

Vilnius, 2023

DUOMENYS. DUOMENŲ TYRYBA

Tikslas

Supažindinti būsimuosius mokytojus su duomenų tyrybos pagrindiniais etapais ir svarbiausiais principais bei padėti pasirengti šios srities mokomosioms veikloms klasėje.

Ryšys su bendrosiomis programomis

(Remiantis atnaujintomis Bendrosiomis programomis)

Informatika

Duomenų tyryba ir informacija

C2. Tyrinėja duomenis ir atlieka veiksmus su jais

7–8 klasė: Tyrinėja duomenų sąryšius, pasitelkia skaitmenines technologijas, statistiką.

III–IV gimnazijos klasė: Tyrinėja ir apibendrina viešai prieinamus ar automatizuotai renkamus duomenis ir išgauna reikalingą informaciją

Matematika

Duomenys ir tikimybės

8, 9–10 klasė ir I–II gimnazijos klasė

Duomenys ir tikimybės. Duomenų interpretavimas

III–IV gimnazijos klasė

Duomenys ir tikimybės. Įvadas į taikomąją duomenų analizę

Ugdomi įgūdžiai

- Duomenų laikymo ir perdavimo kompiuteriuose principų suvokimo.
- Duomenų rinkimo, apdorojimo ir analizės principų suvokimo ir praktinio taikymo.
- Duomenų vizualizavimo pagrindų.

Mokinių amžius: 7–12 klasės

TEORINIS PAGRINDIMAS BŪSIMIESIEMS MOKYTOJAMS

Duomenimis vadinami visi stebimi, renkami ir kaupiami artefaktai, skirti kuriam nors uždaviniui ar problemai spręsti. Tvarkydami, apdorojami duomenis, kuriame informaciją. Šiuolaikiniame pasaulyje duomenys yra didelė vertybė, nes jais remiantis sprendžiamos įvairiausios realaus gyvenimo problemos. Duomenų svarbą lėmė spartus technologinių priemonių ir metodų tobulėjimas, patogios, visiems prieinamos technologijos, kurios leidžia automatizuotai rinkti, kaupti, rūšiuoti ir kitaip apdoroti duomenis. Vystosi atskira tarpdisciplininė mokslo sritis – duomenų mokslas.

Duomenų tyrinėjimo informatikos mokomojo dalyko temos, skirtos duomenų rinkimui, apdorojimui, analizei, vizualizavimui, yra glaudžiai susijusios su tuo, ko mokoma matematikos pamokose (statistikos, tikimybių teorijos konceptai). Informatikos dalies susiejimas su matematikos mokymu yra naudingas integravimo pavyzdys, padedantis pamatyti statistikos ir skaičiuoklės programinės įrangos taikymo naudingumą dirbant su dideliais duomenų kiekiais. Duomenys analizuojami įvairiose srityse, taigi mokytojams atsiveria galimybės su mokiniais įgyvendinti integruotus ne tik informatikos ir matematikos, bet ir kitų dalykų projektus, realizuoti STEM ugdymo principus.

Pateikiame keletą informatikos srities sąvokų, susijusių su duomenų apdorojimu, kurios padės plėtoti mokytojų ir mokinių duomenų tyrybos srities žodyną.

Pagrindinės informatikos dalyko sąvokos

Duomenys – kompiuterio apdorojami objektai – visa, kas laikoma kompiuterinėse laikmenose. Duomenimis išreiškiama apdorojama informacija.

Duomenų tyryba – duomenų apdorojimo, dėsningumų paieškos, analizės, vizualizavimo ir interpretavimo procesų visuma.

Duomenų bazė – duomenų rinkinys, susistemintas ir sutvarkytas taip, kad juo būtų galima patogiai naudotis.

Plokščiasis failas – grynojo teksto failas, sudarytas iš eilutėmis vaizduojamų įrašų, kurių laukai skiriami tabuliavimo ženklais arba kitokiais skirtukais. Paprasčiausias plokščiojo failo pavyzdys – asmenvardžių su kokiais nors duomenimis apie asmenis, pavyzdžiui, adresais, sąrašas.

CSV formatas – duomenų bazės įrašymo tekstiniu formatu (plokščiuoju failu) būdas, kai duomenų laukai skiriami kableliais. Naują įrašą atitinka nauja eilutė. CSV formatas dažnai naudojamas atvirų duomenų saugyklose, adresų knygos duomenims, programos nuostatoms ir kitiems duomenims įrašyti.

Skaičiuoklė – programa, skirta dinaminėms lentelėms rengti ir jose esantiems duomenims apdoroti: atlikti įvairaus sudėtingumo skaičiavimus, grupuoti, rikiuoti, filtruoti, vaizduoti grafiškai.

Lentelė – duomenų pateikimo būdas, kai duomenys išdėstomi stulpeliais ir eilutėmis.

Duomenų filtravimas – duomenų atranka pagal nurodytas sąlygas. Atskiras *rūšiavimo* atvejis, kai objektus vaizduojantys įrašai skirstomi į dvi grupes: atitinkančius filtravimo sąlygą ir jos neatitinkančius.

Diagrama – grafiniu būdu pavaizduoti duomenys. Pagal pavidalą diagrama gali būti stulpelinė, juostinė, taškinė, linijinė, skritulinė, stačiakampė ir pan.

Įvadas į duomenų tyrybą

Skaitmeniniame pasaulyje su duomenimis susiduriame nuolat. Socialinių tinklų programėlės renka duomenis apie mūsų publikacijas, patinkančius įrašus, dalyvių profilių ir naujienų peržiūrą. Internetinės parduotuvės fiksuoja duomenis apie mūsų peržiūrėtas prekes. Reklamos agentūros renka duomenis apie tai, ką spustelime tam tikros svetainės tinklalapyje. Išmanieji įrenginiai skaičiuoja mūsų žingsnius, širdies ritmą, matuoja kitus aktyvumo ir poilsio parametrus.

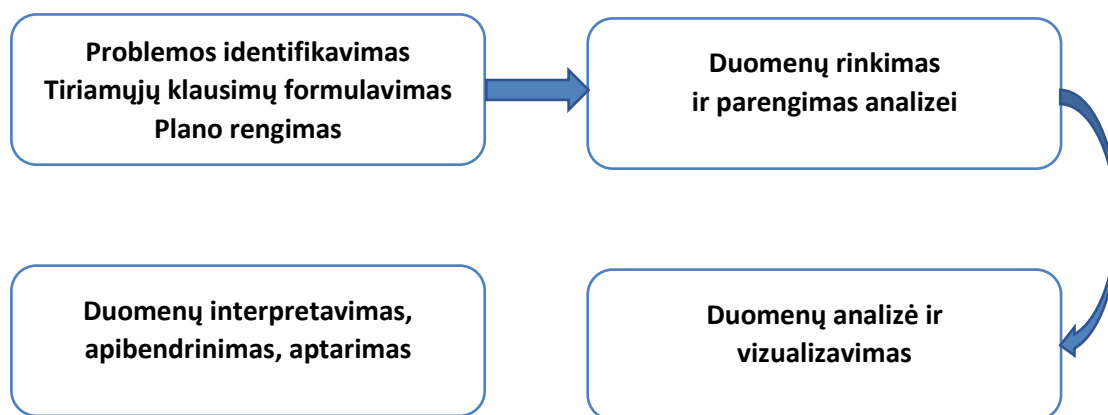
Kaip ir kur laikomi visi šie duomenys? Iš pradžių juos gali rinkti kompiuterio programa, kurios kintamieji laiko duomenis atmintyje. Tačiau toliau duomenis reikia saugoti pastoviojoje atmintyje, nes duomenys turi būti prieinami ir tada, kai programa baigs darbą. Siekiama, kad duomenys būtų saugomi taip, kad juos būtų galima patogiau apdoroti, tirti.

Duomenų saugojimui naudojami įvairūs formatai: nuo paprasčiausių tekstinių failų iki sudėtingų duomenų bazių.

Duomenų tyrybos temos mokykloje padeda pažinti tyrimų metodus, kelti hipotezes ir jas tikrinti, integruojant įvairių dalykų turinį mokytis automatizuoti duomenų apdorojimo procesus, komunikuoti gautus rezultatus.

Duomenų tyrybos procesai

Pagrindinius duomenų tyrybos procesus galime apibūdinti tokia schema.



Problemos identifikavimas, tiriamųjų klausimų formulavimas ir plano rengimas

Tiriamąją problemą siūloma rinktis siejant ją su mokinių interesais, smalsumu, mokiniams aktualiomis situacijomis, temomis, kurių mokomasi per įvairius dalykus mokykloje, atitinkamos klasės matematikos nagrinėjamais konceptais iš duomenų ir tikimybių srities.

Tam kad išspręstume tyrimo problemą, turime suformuluoti ir atsakyti į tiriamuosius klausimus. Tai gali būti klausimai apie duomenų pokyčius, skirtumus tarp tiriamųjų grupių, ryšiai tarp skaitinių kintamųjų ir (arba) klausimai, kuriems atsakyti reikia integruoti daugiau kaip vieną analizės metodą ar priemonę.

Duomenų rinkimas ir parengimas analizei

Duomenys gali būti renkami įvairiais būdais ir iš įvairių šaltinių, pavyzdžiui: iš atvirųjų duomenų šaltinių, kompiuterio programų automatiškai įrašomų duomenų, fiksuojant stebėjimus, matavimus, apklausos būdu.

Šiame etape pasirenkamas duomenų rinkimo būdas, planuojami tinkami tyrimai, atsižvelgiant į iškeltus klausimus ir kontekstą.

Duomenų rinkimo procese susiduriama su tuo, kad priimama daug sprendimų, kurie turės įtakos duomenų rinkimo kokybei, kiekiui, rinkimo spartai, duomenų prieinamumui.

Surinkus duomenis, jie parengiami tolesnei analizei: išvalomi, sutvarkomi, įskaitant neteisingų duomenų identifikavimą, sutvarkymą ir (arba) šalinimą, dalinių rinkinių sudarymą, filtravimą, naujų kintamųjų kūrimą.

Duomenų analizė ir vizualizavimas

Skaiciuojamos įvairios duomenų charakteristikos, taikomos duomenų išrinkimo ir analizės priemonės. Ieškoma dėsningumų, šablonų duomenyse, koreliacijų tarp kintamųjų, skirtumų tarp grupių, sudaromi modeliai. Taikomos įvairios analizę automatizuojančios priemonės, pavyzdžiui, skaičiuoklė. Tikrinamos iškeltos hipotezės.

Duomenys vizualizuojami taikant įvairių tipų diagramas, lenteles ir kitas priemones, tinkamas konkrečiam atvejui.

Duomenų interpretavimas ir komunikavimas

Formuluojami atsakymai į iškeltus tiriamuosius klausimus.

Nagrinėjamos duomenų tendencijos, prieštaravimai. Atkreipiamas dėmesys į tai, kokias išvadas patvirtina duomenys. Interpretuojami skaičiavimų rezultatai, pavyzdžiui, koreliacijos koeficientai.

Duomenų analizės rezultatai pateikiami rengiant pristatymus skirtingoms bendruomenės grupėms, reaguojant į skirtingus argumentus ar duomenų interpretacijas.

Praktinė užduotis. Darbas porose

Raskite vieną iš šaltinių (pvz., naujienų straipsnį), kuriame būtų pristatomi tam tikrų tyrimų duomenų analizės rezultatai ir vaizduojami grafiškai. Dirbkite poroje ir aptarkite klausimus:

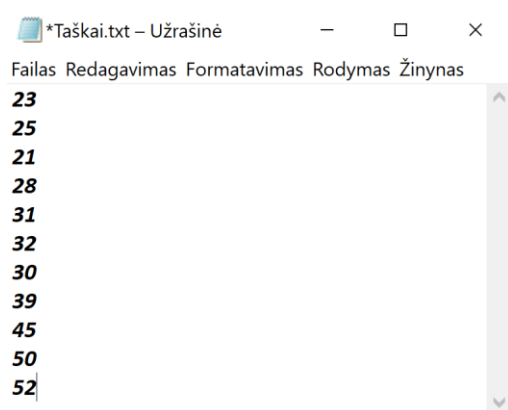
- Kaip Jūs manote, kaip buvo surinkti duomenys?
- Kaip pateikiami rezultatai?
- Kuris diagramų tipas parinktas duomenims vaizduoti grafiškai?
- Kodėl, Jūsų nuomone, pasirinktas būtent toks diagramos tipas?
- Aptarkite, kaip interpretuoti grafiškai pateiktus duomenis.
- Patikrinkite, ar yra tokio diagramos tipo atitikmuo Jūsų naudojamoje skaičiuoklėje (pvz., „MS Excel“, „LibreOffice Calc“).

DUOMENŲ PATEIKIMAS

Tekstiniai duomenų failai

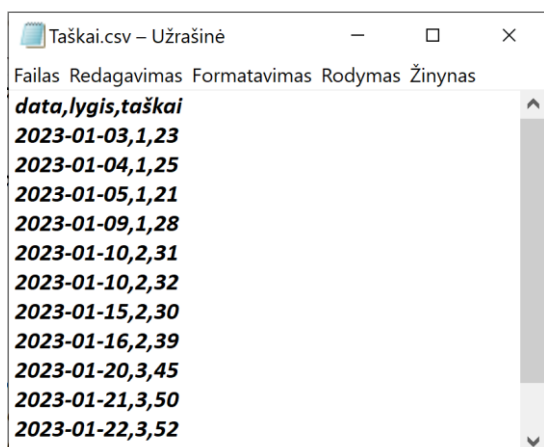
Dauguma kompiuterio programų fiksuoja duomenis apie programų naudojimą, naudotojo pasirinkimus, pasiekimus ir kt. veiksmus tekstiniuose failuose, kurie laikomi kompiuterio diske. Pavyzdžiui, kompiuteriniai žaidimai fiksuoja pasiekiamus rezultatus, kad būtų galima stebėti pažangą.

Žaidimų rezultatai gali būti laikomi paprasčiausiame tekstiniame faile „taškai.txt“, kuriame kiekvienoje eilutėje įrašyti žaidimo taškai. Tokius failus galima peržiūrėti grynojo teksto rengykle, pvz., užrašine.



Jei reikėtų saugoti daugiau duomenų, t. y., ne tik taškus, bet ir žaidimo datą, pasiektą žaidimo lygį ir taškus, reikėtų keletą stulpelių, t. y., kiekvienoje failo eilutėje reikėtų saugoti keletą susijusių duomenų elementų. Tokio tipo duomenys sudaro lentelę. Kiekviena eilutė yra lentelės eilutė, o kiekviena eilutė turi keletą stulpelių.

Tokius lentelių duomenų pateikimo failus vadiname *plokščiaisiais*. Labiausiai paplitęs plokščiųjų failų – lentelių duomenų laikymo tekstiniuose failuose formatas – CSV (kableliais atskirtos reikšmės). Tokiame faile toje pačioje eilutėje esančios reikšmės skiriamos kableliais ir laikomos faile, pvz., „Taškai.csv“.



Pirmoje failo eilutėje apibrėžiami stulpeliai: data, lygis ir taškai. Kitose eilutėse pateikiamos duomenų reikšmės, kuriose pirma įrašoma data, po to lygis, tada rezultatas, o visos reikšmės skiriamos kableliais.

CSV failai – paplitęs formatas (juo dažnai pateikiami duomenys atvirose duomenų saugyklose), todėl yra daug priemonių jam skaityti ir rašyti. Gali būti naudojami ir kitokie skirtukai tos pačios eilutės skirtingoms duomenų reikšmėms atskirti, pvz., tabuliacijos ženklai (failų prievardis TAB arba TSV).

CSV ir kiti plokštieji failai duomenims laikyti turi tam tikrų trūkumų – jiems reikia papildomų priemonių duomenims apdoroti.

CSV failus (taip pat kitokius skirtukus naudojančius failus) gali skaityti ir įrašyti skaičiuoklės. Skaičiuoklėse galima apdoroti CSV failuose pateikiamus duomenis, prieš tai importavus juos į skaičiuoklės formatą. Programuotojai taip pat naudojami programavimo kalbų bibliotekomis, skirtomis CSV failams skaityti ir rašyti iš jų kuriamų programų.

Skaičiuoklės

Skaičiuoklė – tai programa, skirta dinaminėms lentelėms rengti ir jose esantiems duomenims apdoroti: atlikti įvairaus sudėtingumo skaičiavimus, grupuoti, rikiuoti, filtruoti, vaizduoti grafiškai.

Skaičiuoklės paprastai gali atverti įvairių formatų (pvz., CSV, TSV, XLS) failus ir įrašyti duomenis atgal tuo pačiu formatu.

Pastaba. Skiriant duomenų reikšmes kableliais, susiduriame su nevienareikšmiškumu, nes Lietuvoje ir kitose Europos šalyse kablelis naudojamas dešimtainės trupmenos sveikajai ir trupmeninei dalims atskirti. Todėl skaičiuoklės, eksportuodamos lentelę į CSV failą, gali skirti reikšmes kabliataškiais.

Čia parodytas anksčiau aptartas failas „Taškai.csv“, importuotas į „MS Excel“ skaičiuoklę.

	A	B	C	
1	data	lygis	taškai	
2	2023.01.03	1	23	
3	2023.01.04	1	25	
4	2023.01.05	1	21	
5	2023.01.09	1	28	
6	2023.01.10	2	31	
7	2023.01.10	2	32	
8	2023.01.15	2	30	
9	2023.01.16	2	39	
10	2023.01.20	3	45	
11	2023.01.21	3	50	
12	2023.01.22	3	52	

Tam, kad kableliais ar kitais ženklais atskirtas reikšmes skaičiuoklė konvertuotų į savo formatą, atveriant tokį failą pateikiamas importo vediklis (pavyzdyje – „MS Excel“ skaičiuoklės importo vediklis). Jame pasirenkama importuojamų duomenų koduotė, atliekama preliminarinių duomenų peržiūra.

Teksto importavimo vediklis - 1 veiksmas iš 3

Teksto vediklis nustatė, kad jūsų duomenų tipas yra Atskirti.

Jei ši informacija teisinga, pasirinkite Pirmyn. Kitu atveju pasirinkite duomenų tipą, kuris tiksliausiai apibūdina jūsų duomenis.

Pirminių duomenų tipas

Pasirinkite failo tipą, kuris tiksliausiai apibūdina jūsų duomenis:

☒ Atskirti - Laukus skiria tokie simboliai kaip kableliai ar tabuliacijos žymės.

☐ Eiksuoto pločio - Laukai susilygiuoti į stulpelius, tarp kiekvieno lauko paliktas tarpas.

Pradėti importuoti nuo eilutės: 1 Failo kilmė: 65001 : Unicode (UTF-8)

☐ Mano duomenys turi antraštes.

Failo C:\Users\tatja\Documents\Projektai\2021-Skaitmenini... (Absolventų skaičius pagal studijų grupes - ST ir ICT) peržiūra.

1	ST Metai, kada gautas diplomas	ST Padalinio savivaldybės ID	ST Padalinio savivaldybės	ST Padali
2	2022	2	Alytaus m. sav.	3911
3	2022	2	Alytaus m. sav.	3911
4	2022	2	Alytaus m. sav.	3911
5	2022	2	Alytaus m. sav.	3911
6	2022	2	Alytaus m. sav.	3911
7	2022	2	Alytaus m. sav.	3911

Atšaukti < Atgal Pirmyn > Baigti

Kitame žingsnyje pasirenkamas laukus skiriantis ženklas, pvz., tabuliacijos (TAB, TSV), kablelis (CSV) ir kt.

Teksto importavimo vediklis - 2 veiksmas iš 3

Šiame ekrane galite nustatyti savo duomenų skyriklius. Apačioje, peržiūroje, pamatysite, kaip tai paveiks jūsų tekstą.

Skyrikliai

☒ Tabuliacijos ženklas

☐ Kablelis

☐ Tarpas

☐ Kita:

☐ Iš eilės einančius skyriklius laikyti vienu

Teksto kvalifikatorius: "

Duomenų peržiūra

ST Metai, kada gautas diplomas	ST Padalinio savivaldybės ID	ST Padalinio savivaldybės	ST Padalinio
2022	2	Alytaus m. sav.	3911
2022	2	Alytaus m. sav.	3911
2022	2	Alytaus m. sav.	3911
2022	2	Alytaus m. sav.	3911
2022	2	Alytaus m. sav.	3911
2022	2	Alytaus m. sav.	3911

Atšaukti < Atgal Pirmyn > Baigti

Paskutiniame žingsnyje pasirenkamas duomenų formatas.

Praktinė užduotis

Išnagrinėkite pasirinktą atvirųjų duomenų portalą ir Jūsų pasirinktai temai parsiųskite duomenų lentelę CSV formatu. Importuokite duomenis į skaičiuoklę ir parenkite juos tolesniam apdorojimui. Praktiškai išbandę procesą, aptarkite, į ką reikėtų atkreipti mokinių dėmesį.

Keletas atvirųjų duomenų portalų pavyzdžių duomenims pasirinkti:

- Lietuvos atvirųjų duomenų portalas, <https://data.gov.lt>
- Oficialus Europos duomenų portalas, <https://data.europa.eu/euodp/lt>
- Mūsų pasaulis duomenyse. *Our world in data*, <https://ourworldindata.org/>

Aptarkite kitas žinomas atvirųjų duomenų saugyklas.

DUOMENŲ ANALIZĖ IR VIZUALIZAVIMAS

Pagrindinės duomenų apdorojimo priemonės skaičiuoklėje

Prisiminkite su mokiniais matematikos statistikos temų pamokose nagrinėjamas sąvokas: imtis, kintamasis, vidurkis, mediana, kvartilis, koreliacija, koreliacijos koeficientas, Pirsono koreliacijos koeficientas, koreliacija ir priežastingumas ir kt.

Skaičiuoklės turi priemonių daugumos pagrindinių statistinių rodiklių skaičiavimams automatizuoti.

Išnagrinėkite skaičiuoklės funkcijas ir dešinėje parašykite kiekvienos jų trumpą apibūdinimą bei taikymo pavyzdį Jūsų naudojamoje skaičiuoklės versijoje.

<i>Funkcija</i>	<i>Apibūdinimas</i>	<i>Taikymo pavyzdys</i>
AVERAGE		
MEDIAN		
MIN		
MAX		
SUM		
COUNT		
COUNTA		
COUNTIF		
SUMIF		
CORREL		

Kurios iš šių funkcijų tinka duomenims filtruoti?

Duomenų vizualizavimas – tai ne tik vaizdinis duomenų pateikimo būdas, padedantis komunikuoti duomenų analizės rezultatus, bet ir svarbi duomenų analizės priemonė, leidžianti ieškoti dėsningumų duomenyse.

Diagramų tipų analizė. Praktinė užduotis

Išanalizuokite vieną diagramos tipą skaičiuoklėje ir parenkite trumpą pristatymą grupei, kuriame būtų apžvelgiama:

- Diagramos pavadinimas ir bendras apibūdinimas.
- Duomenų lentelės pavyzdys, tinkamas tokio tipo diagramai braižyti.
- Nubraižyta nurodyto tipo diagrama skaičiuoklėje.
- Galimos diagramos modifikacijos (potipiai, vaizdavimo būdų pakeitimai).
- Apibūdinti pagrindiniai atvejai, kada šis diagramos tipas labiausiai tinka duomenims vizualizuoti.

Diagramų pavadinimų sąrašas diagramų tipų analizei:

Nr.	Diagramos tipo pavadinimas
1	Juostinė diagrama (angl. <i>bar chart</i>)
2	Skritulinė diagrama (angl. <i>pie chart</i>)
3	Žiedinė diagrama (angl. <i>doughnut chart</i>)
4	Spindulinė diagrama (angl. <i>radar chart</i>)
5	Taškinė diagrama (angl. <i>xy (scatter) chart</i>)
6	Burbulinė diagrama (angl. <i>bubble chart</i>)
7	Linijinė diagrama (angl. <i>line chart</i>)
8	Sritinė diagrama (angl. <i>area chart</i>)
9	Kombinuotoji diagrama (angl. <i>combo chart</i>)
10	Sudurtinė stulpelinė diagrama (angl. <i>stacked column chart</i>)
11	Procentinė stulpelinė diagrama (angl. <i>100% stacked column chart</i>)
12	Akcijų diagrama (angl. <i>stock chart</i>)

Diagramų tipų apibrėžčių ieškokite kompiuterijos žodynuose (www.raštija.lt).

Duomenis pasirinkto tipo diagramai nubraižyti pasirinkite iš anksčiau analizuotų atvirųjų duomenų šaltinių.

Duomenų vizualizavimas stačiakampe diagrama

Informatyvus duomenų pasiskirstymo tyrinėjimo ir vizualizavimo būdas, dažnai naudojamas įvairiose mokslo publikacijose, – stačiakampės diagramos (angl. *box plot*). Tokias diagramas galima kurti analizuojant duomenis specializuotuose statistikos paketuose, pavyzdžiui, R, IBM SPSS ir kt. Naujesnės skaičiuoklių versijos turi galimybės braižyti tokias diagramas. Šio tipo diagramų taikymas duomenims vizualizuoti numatytas informatikos ir matematikos Bendrosiose programose.

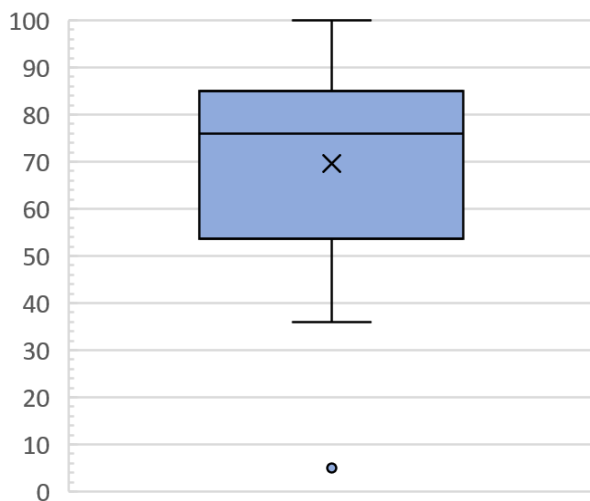
Praktinė užduotis

Tarkime, turime 9 klasės 50 mokinių už semestrą surinktus taškus. Kiekvienas mokinys identifikuojamas numeriu 01, 02, 03 ir t. t., o jo taškai 9 klasėje surašyti stulpeliuose „9 klasė“.

Paveiksle matome stačiakampę diagramą, nubraižytą skaičiuokle „MS Excel“.

ID	9 klasė
01	5
02	58
03	91
04	76
05	70
06	99
07	53
08	87
09	50
10	41
11	36
12	85
13	68
14	90

9 klasės mokinių taškų pasiskirstymas



1. Prisiminkite *kintamojo* sąvoką. Kiek kintamųjų yra lentelėje? Kiek kintamųjų pavaizduota stačiakampe diagrama?
2. Paeksperimentuokite su duomenimis ir jų vaizdavimu stačiakampe diagrama praktiškai.
3. Naudodamiesi įvairiais literatūros šaltiniais (statistikos vadovėliais, internete rasta šaltiniais) išnagrinėkite, ką ši diagrama „sako“ apie duomenis.
4. Į paveikslo dešinėje esančius tuščius stačiakampius įrašykite atitinkamus pavadinimus, pasirinkę iš sąrašo:

Viršutinė riba

Apatinė riba

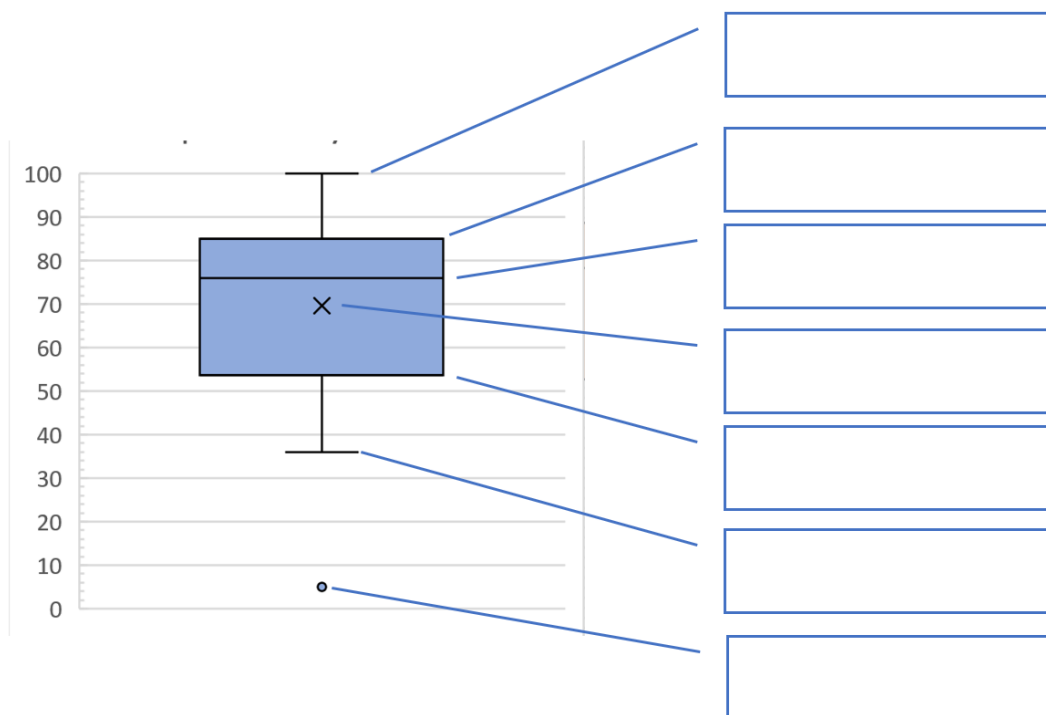
Vidurkis

Mediana

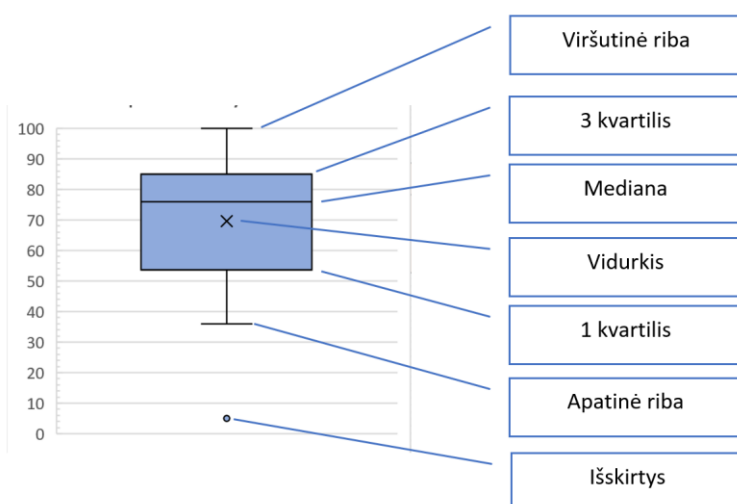
1 kvartilis

3 kvartilis

Išskirtys (angl. *outliers*)

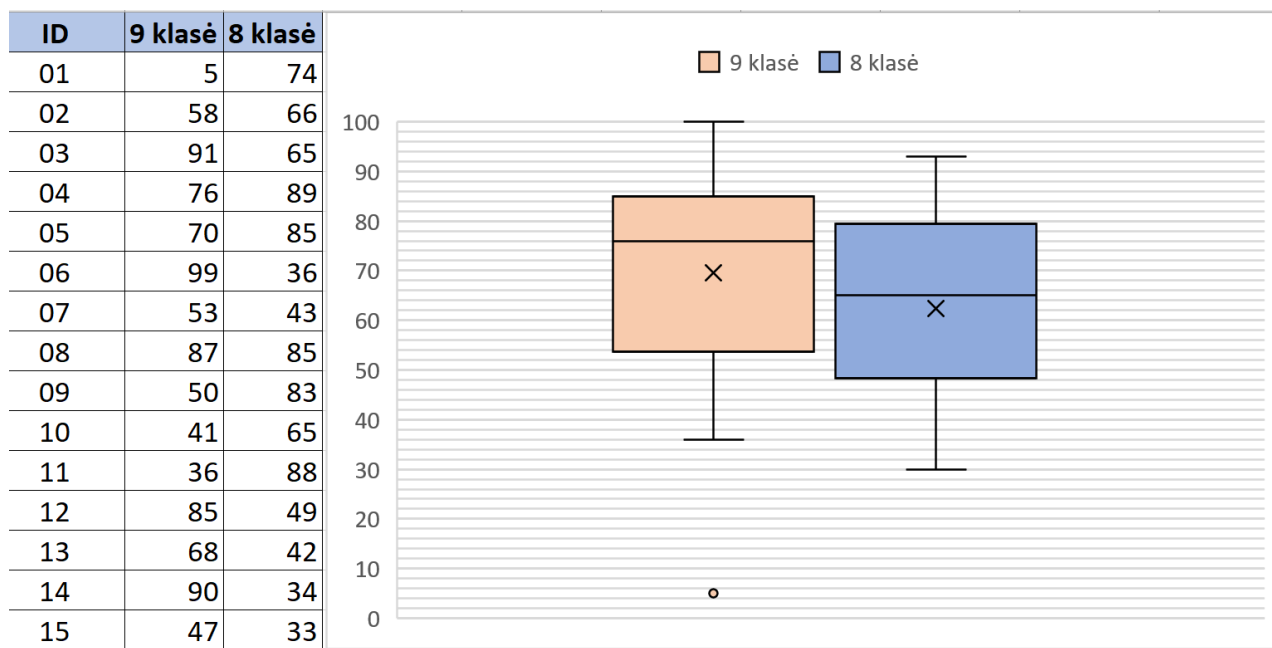


<Atsakymas:>



Stačiakampėmis diagramomis galima vaizduoti ir daugiau kaip vieną kintamąjį, tokiu būdu palyginant dviejų grupių duomenis arba tos pačios grupės rodiklių pokytį.

Pateikiamas pavyzdys, papildytas 8 klasės duomenimis. Kiekvienas mokinys identifikuojamas numeriu 01, 02, 03 ir t. t., o jo taškai 9 ir 8 klasėje surašyti stulpeliuose „9 klasė“ ir „8 klasė“. Stačiakampe diagrama vaizduojami 9 ir 8 klasių rezultatai.



5. Porose aptarkite, kokias išvadas galime daryti iš daugiau kaip vieno kintamojo vaizdavimo stačiakampe diagrama.
6. Išanalizuokite matematikos bendrąją programą ir raskite konceptus, kurių mokiniai gali mokytis sudarydami stačiakampes diagramas (kuriose klasėse?).

Stačiakampių diagramų sudarymo principai

Naudojama išsibarstymo charakteristika, pagrįsta skirtumu tarp 3-jo ir 1-jo kvartilų.
Tarpkvartilinio diapazono metodas:

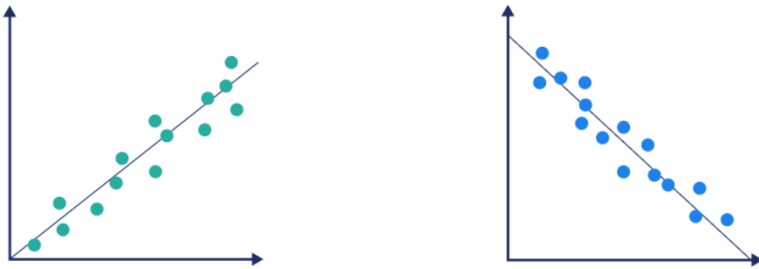
1. Duomenys surikiuojami didėjimo tvarka.
2. Identifikuojamas pirmas kvartilis (Q1), mediana ir trečias kvartilis (Q3).
3. Skaičiuojamas tarpkvartilinis diapazonas $IQR = Q3 - Q1$.
4. Skaičiuojama viršutinė riba: $Q3 + (1,5 \times IQR)$.
5. Skaičiuojama apatinė riba: $Q1 - (1,5 \times IQR)$.
6. Mažesnės už apatinę ribą ir didesnės už viršutinę ribą reikšmės laikomos išskirtimis.

Išskirtys (angl. *outliers*) dažnai byloja apie tai, kad imtyje yra „nereguliarių“ atvejų arba „klaidų“.

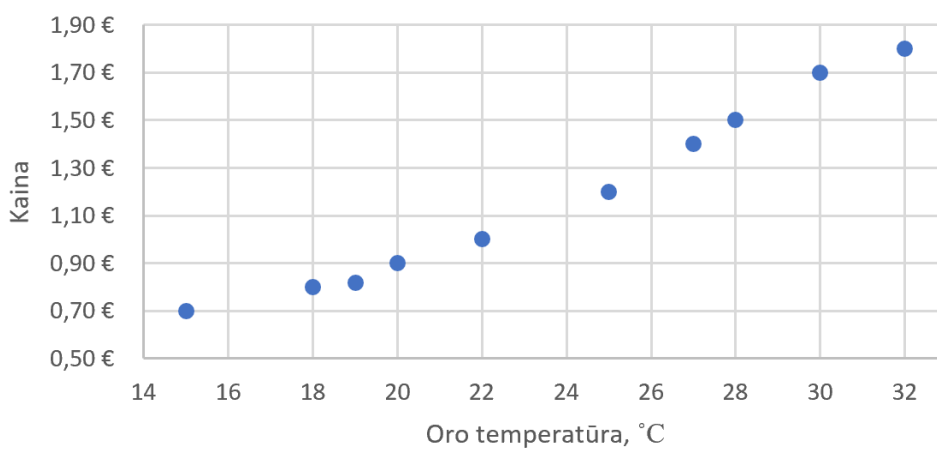
Ryšių tarp kintamųjų vizualizavimas

Vienas iš svarbių duomenų analizės tikslų – rasti koreliaciją, statistinį ryšį tarp dviejų kintamųjų. Koreliacija gali būti teigiama, neigiama arba jos gali nebūti. Dažniausias būdas dviejų kintamųjų koreliacijai vizualizuoti yra taškinė diagrama (arba jos variantas – burbulinė diagrama).

Paveiksle schematiškai pavaizduota stipri teigiama tiesinė koreliacija (kairėje) ir stipri neigiama (dešinėje).

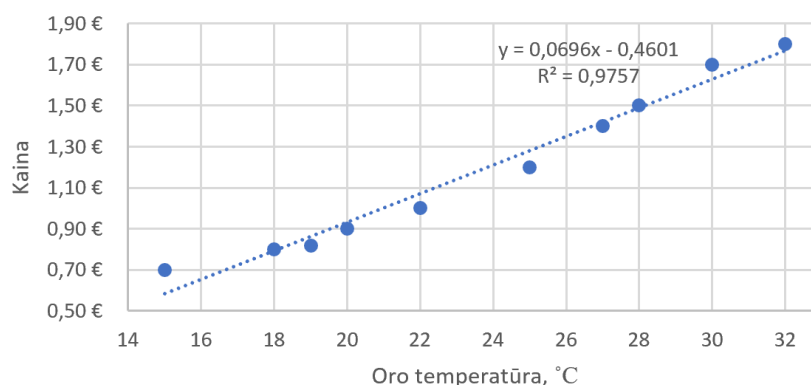
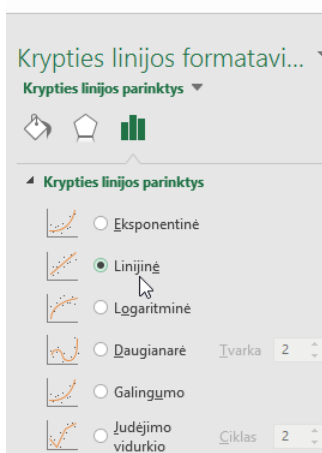


Skaičiuoklės priemonėmis galima vizualizuoti ryšį tarp dviejų kintamųjų X ir Y taikant taškinę diagramą. Pavyzdžiui, taip pavaizduotas oro temperatūros ir ledų kainos ryšys:



Praktinės užduotys

1. Skaičiuoklės priemonės (krypties linija, angl. *trend line*) leidžia rasti duomenis aprašantį modelį (tiesinį, eksponentinį, logaritminį ir pan.). Pateikto oro temperatūros ir ledų kainos pavyzdžio tiesinis modelis, jo lygtis ir parametras R^2 pavaizduoti dešinėje.



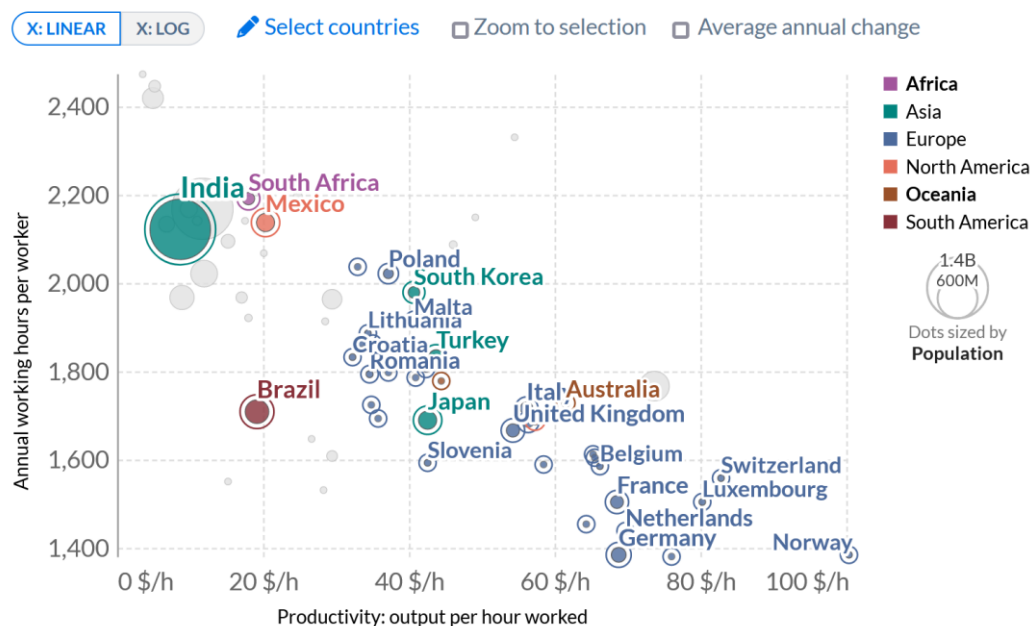
Aptarkite klausimus:

- Kaip galėtume apibūdinti koreliaciją tarp oro temperatūros ir ledų kainos?

- Ką šiame pavyzdyje reiškia R^2 ?
- Kokios ledų kainos galima tikėtis, kai oro temperatūra yra 24° , 34° ? Ar patikima tokia prognozė?

2. Pateiktoje burbulinėje diagramoje vaizduojamas darbo našumas įvairiose šalyse, matuojamas kaip bendrasis vidaus produktas (BVP), tenkantis vienai darbo valandai (2019 m. duomenimis). BVP koreguojamas atsižvelgiant į infliaciją ir pragyvenimo išlaidų skirtumus įvairiose šalyse. Šaltinis (su interaktyviomis diagramomis):

<https://ourworldindata.org/working-hours>



Išnagrinėkite diagramą ir pasirinkite vieną iš atsakymų variantų, kuris geriausiai apibūdina ryšį tarp darbo našumo ir vidutinio dirbtų valandų skaičiaus:

- Tarp produktyvumo ir vidutinio darbo valandų skaičiaus yra teigiama koreliacija.
- Tarp produktyvumo ir vidutinio darbo valandų skaičiaus yra neigiama koreliacija.
- Tarp produktyvumo ir vidutinio darbo valandų skaičiaus koreliacijos nėra.

<Teisingas atsakymas: B>

Užduotis

Parenkite integruoto projekto mokiniams scenarijų, įtraukdami visus pagrindinius duomenų tyrimo proceso etapus ir duomenų apdorojimo bei automatizavimo priemones. Parengtą idėją pristatykite savo studijų grupėje.

Patarimai:

- Suteikite mokiniams galimybę rinkti duomenis, tvarkyti ir vaizduoti duomenis įvairiuose kontekstuose.
- Naudokite kasdienes galimybes duomenims tyrinėti ir šios srities žodynui pažinti.
- Pasiūlykite mokinių interesus atitinkančias temas, kurias nagrinėdami jie galėtų rinkti ir analizuoti duomenis.

Šaltiniai

1. Dagienė V., Grigas G., Jevsikova T. Enciklopedinis kompiuterijos žodynas.
<http://www.raštija.lt>; <http://www.ims.mii.lt/EKŽ/>
2. Priešmokyklinio, pradinio, pagrindinio ir vidurinio ugdymo bendrųjų programų 21 priedas. Informatikos bendroji programa, 2022.
https://www.emokykla.lt/upload/EMOKYKLA/BP/2022-10-10/PATVIRTINTA_Aurelija/21_Informatikos%2BBP%2B2022-09-30.pdf
3. Priešmokyklinio, pradinio, pagrindinio ir vidurinio ugdymo bendrųjų programų 20 priedas. Matematikos bendroji programa, 2022.
https://www.emokykla.lt/upload/EMOKYKLA/BP/2022-10-10/20_Matematikos%20BP%202022-09-08.pdf
4. Khan Academy. Unit: Data analysis. <https://www.khanacademy.org/computing/ap-computer-science-principles/data-analysis-101#data-tools>
5. Our world in data. <https://ourworldindata.org/>
6. Stanford graduate school of education. Data science K-10 big ideas. Youcubed, 2021.
https://www.youcubed.org/data-big-ideas/#grade_8-10