

NOTES ON BASIS OF CDA PPT INTRODUCTION TO R PROGRAMMING 27 SLIDES

CHEAT SHEET - Read PPT once

Introduction to Computational Data Analytics Cheat Sheet:

Computational Data Analytics:

- Field for depth and specialization in data science: ML, deep learning, natural language, AI, visualization, databases, high-performance computing, etc.
- Examples of computational thinking: chess strategy, map reading, task organization.
- Steps of Computational Thinking: Abstraction, Automation, Analysis.
- Principals of Computational Thinking: Decomposition, pattern recognition, abstraction, algorithms.
- Benefits of computational thinking: Real-world problem solving, simplifying complex problems.

Computational Skills:

- Ability to perform basic arithmetic quickly and accurately using mental methods, paper-and-pencil, or calculator.

Types of Computation:

- Models of computation: Sequential models, functional models, concurrent models.
- Purpose of Computational: Intelligent health information analysis for disease treatment guidance.

Computational Analytics:

- Enables scientific discovery through algorithms identifying patterns, anomalies, testing hypotheses, creating models.
- Computational Data Science: Combines statistics, computer science, math, ML for trend identification, prediction, problem-solving.
- Uses algorithms, data structures for storage, manipulation, visualization, learning from large datasets.

Data Analytics:

- Process of examining data sets to find trends and draw conclusions.
- Increasingly aided by specialized systems and software.

Introduction to R Programming:

- Open-source programming language for statistical software and data analysis.
- Widely used, available on Windows, Linux, macOS.
- Command-line interface.

Why R Programming Language?:

- Leading tool for machine learning, statistics, data analysis.
- Platform-independent, open-source, integrates with other languages.
- Growing community of users, high demand in Data Science job market.

Features of R Programming Language:

- Statistical Features: Basic statistics, static graphics, probability distributions, data analysis.
- Programming Features: Abundance of packages (CRAN), distributed computing.

Programming in R:

- Similar syntax to other languages, easy to learn and code.
- Write and save programs with .r extension, run with "R filename.r" command.

Advantages of R:

- Comprehensive statistical analysis package.
- Open-source, cross-platform, active community.

Disadvantages of R:

- Some packages may have lower quality.
- Memory management challenges.
- Limited formal support.

Applications of R:

- Data Science, quantitative analysis, finance, tech giants like Google, Facebook, etc.

R and Data Science:

- R and Python play major roles in data science.
- Data science involves identifying, representing, extracting meaningful information for business logic.

Tools for Data Science:

- R, Python, SQL, SAS, Tableau, MATLAB.
- R and Python are most used.

R vs Python in Data Science:

- Specialities: R has advanced statistical techniques, Python is good for web development.
- Functionalities: R has inbuilt data analysis functions, Python relies on packages.
- Domains: R excels in data visualization, Python in deep learning.
- Packages: R has numerous packages, Python relies on Scikit-learn, Pandas.

Remember, this cheat sheet provides a concise overview of the key concepts covered in the introduction to Computational Data Analytics, including aspects of Computational Thinking, R Programming, and the comparison between R and Python for Data Science.

NOTES

Introduction to Computational Data Analytics - Notes:

Computational Data Analytics:

- Field for specialization in data science (ML, deep learning, natural language, AI, etc.) building on interdisciplinary core curriculum.
- Computational thinking examples: chess strategy, map reading, task organization.
- Steps of Computational Thinking: Abstraction, Automation, Analysis.
- Principals of Computational Thinking: Decomposition, pattern recognition, abstraction, algorithms.
- Computational thinking benefits: Real-world problem solving, breaking down complex problems.

Computational Skills:

- Ability to perform basic arithmetic accurately and quickly using mental methods, calculators, etc.

Types of Computation:

- Models of computation: Sequential, functional, concurrent.
- Purpose of Computational: Intelligent health data analysis for disease treatment guidance.

Computational Analytics:

- Uses algorithms for pattern identification, anomaly detection, hypothesis testing, model creation, uncertainty quantification.
- Computational Data Science: Combines statistics, computer science, math, ML for trend identification, prediction, problem-solving.
- Uses algorithms, data structures for storage, manipulation, visualization, learning from large data sets.

Data Analytics:

- Examining data sets to discover trends, draw conclusions.
- Increasing use of specialized systems and software for data analytics.

Introduction to R Programming:

- Open-source language for statistical software, data analysis.
- Widely used, supports Windows, Linux, macOS.
- Command-line interface.

Why R Programming Language?:

- Leading tool for ML, statistics, data analysis.
- Platform-independent, open-source, integrates with other languages.
- Growing user community, high demand in Data Science job market.

Features of R Programming Language:

- Statistical Features: Basic statistics (mean, mode, median), static graphics, probability distributions, data analysis.
- Programming Features: Abundance of packages (CRAN), distributed computing.

Programming in R:

- Similar syntax to other languages, easy learning.
- Write .r programs, run using "R filename.r" command.

Advantages of R:

- Comprehensive statistical analysis package.
- Open-source, cross-platform, active community.

Disadvantages of R:

- Some package quality issues.
- Memory management challenges.
- Limited formal support.

Applications of R:

- Data Science, quantitative analysis, finance, tech giants like Google, Facebook.

R and Data Science:

- R and Python important in data science.
- Data science: Identify, represent, extract meaningful data for business logic.
- Data scientists use ML, statistics for analysis and decision-making.

Tools for Data Science:

- R, Python, SQL, SAS, Tableau, MATLAB, with R and Python being popular.
- Choosing between R and Python can be confusing for newcomers.

R vs Python in Data Science:

- R has advanced statistical techniques, Python covers common techniques.
- R packages cover various domains, Python's packages like Scikit-learn and Pandas are popular.
- R excels in data visualization, Python better for web development.
- Python better for deep learning and neural networks.
- R has abundant packages but may require more specialization, Python has fewer main packages but easier to use for common tasks.

SUMMARY

Summary: Introduction to Computational Data Analytics

Computational Data Analytics is a specialized field within data science that enables students to delve deeper into topics such as machine learning, deep learning, natural language processing, AI, visualization, databases, and high-performance computing. It encompasses computational thinking, which involves problem formulation, solution expression, and solution execution and evaluation. Principles of computational thinking include decomposition, pattern recognition, abstraction, and algorithms. Computational thinking fosters real-world problem-solving by breaking down complex issues into manageable steps.

Computational skills involve performing basic arithmetic operations rapidly and accurately using mental methods, calculators, and other tools. There are different types of computation models, including sequential, functional, and concurrent, with applications ranging from healthcare data analysis to disease treatment guidance.

Computational Analytics facilitates scientific discoveries through algorithms that identify patterns, anomalies, test hypotheses, create models, and quantify uncertainties. Computational Data Science integrates statistics, computer science, mathematics, and machine learning to identify trends, make predictions, and solve problems. It utilizes algorithms and data structures for manipulating, visualizing, storing, and learning from large datasets.

Data Analytics involves examining datasets to uncover trends and draw conclusions. R, an open-source programming language, is widely used for statistical software and data analysis. It offers statistical features like basic statistics, static graphics, probability distributions, and data analysis tools. R packages are abundant and offer distributed computing capabilities.

Despite some limitations, R has become a comprehensive statistical analysis package, widely used in data science, quantitative analysis, and finance. It is a fundamental tool for various applications, including tech giants like Google and Facebook.

In comparison with Python, R has strengths in advanced statistical techniques and data visualization. While Python covers common techniques, it excels in deep learning and web development. Both R and Python play significant roles in data science, with each offering unique advantages in different domains.

In conclusion, computational data analytics is a dynamic field that combines various disciplines to enable data-driven insights and problem-solving. The choice between R and Python depends on specific requirements and preferences, with both languages contributing to the vibrant landscape of data science tools.

SUMMARY IN POINTS

Summary: Introduction to Computational Data Analytics in Points

1. Computational Data Analytics is a specialized field that offers depth in data science, including ML, deep learning, AI, visualization, and high-performance computing.
2. Computational thinking involves problem formulation, solution expression, and solution execution/evaluation.
3. Principles of Computational Thinking: decomposition, pattern recognition, abstraction, and algorithms.
4. Computational thinking helps break down complex problems into manageable steps for real-world problem solving.
5. Computational skills involve rapid and accurate mental calculations using tools like calculators.
6. Types of Computation: sequential, functional, and concurrent models.
7. Computational models gather, filter, analyze, and present health data for personalized disease treatment.
8. Computational Analytics uses algorithms to identify patterns, anomalies, and quantifies uncertainties in scientific discovery.
9. Computational Data Science combines stats, computer science, math, and ML to identify trends, predict, and solve problems.
10. Data Analytics examines data sets for trends and conclusions, often aided by specialized systems and software.
11. Introduction to R Programming: Open-source language for statistical software and data analysis.

12. R supports command-line interface, available on Windows, Linux, macOS.
13. R is used for machine learning, statistics, and data analysis, with platform independence and integration capabilities.
14. R boasts a vast community, making it popular in the data science job market.
15. R's Statistical Features: Measures of Central Tendency, static graphics, probability distributions, data analysis tools.
16. R Packages provide extensive libraries, including CRAN repository with over 100,000 packages.
17. Distributed Computing in R through packages like ddR and multidplyr.
18. Programming in R is similar to other languages, with options like R Studio, Tinn-R.
19. Advantages of R: Comprehensive statistical package, open-source, cross-platform, collaborative community.
20. Disadvantages of R: Inconsistent package quality, memory management challenges, slower than some languages.
21. Applications of R: Data science, quantitative analysis, finance, used by tech giants.
22. R and Python play major roles in data science, with specific strengths in advanced techniques and deep learning, respectively.
23. Tools for Data Science include R, Python, SQL, SAS, Tableau, MATLAB.
24. Data visualization emphasized; R's ggplot2, ggvis, Python's Bokeh, Matplotlib.
25. Python excels in deep learning with packages like Lasagne, Keras, TensorFlow.
26. Availability of Packages: R has many packages, Python relies on Scikit learn, Pandas for common techniques.

DIAGRAMS - Refer PPT Diagrams also

Creating a visual diagram with text-based representation can be challenging in a text-only format. However, I can provide you with a simplified text-based representation of the diagram you're describing:

...

Introduction to Computational Data Analytics

| Computational Data Analytics |

- | - Depth in data science: ML, deep learning, AI, visualization, databases, etc. |
- | - Examples of computational thinking: chess strategy, maps, task organization |
- | - Steps of Computational Thinking: Abstraction, Automation, Analysis |
- | - Principals of Computational Thinking: Decomposition, Pattern Recognition, Abstraction, Algorithms |
- | - Computational skills: Rapid mental calculations using tools like calculators |
- | - Types of Computation: Sequential, Functional, Concurrent models |
- | - Purpose of Computational: Intelligent health information analysis for disease treatment |

| Computational Analytics |

- | - Algorithms for scientific discovery: pattern identification, anomaly detection |
- | - Combines statistics, CS, math, ML for trend identification, predictions, problem-solving |
- | - Data manipulation, visualization, and learning from large datasets |

| Data Analytics |

- | - Examining data sets for trends and conclusions |
- | - Increasing use of specialized systems and software |

| Introduction - R Programming |

- | - Open-source language for statistical software and data analysis |
- | - Widely used for ML, stats, data analysis on multiple platforms |
- | - Latest cutting-edge tool |

| Why R Programming Language? |

- | - Leading tool for ML, statistics, data analysis |
- | - Platform-independent |
- | - Open-source, integration with other languages |
- | - Vast and growing community |

| Features of R Programming Language |

- | - Statistical Features: Measures of Central Tendency, Static Graphics, Probability Distributions |
- | - Programming Features: Extensive Packages, Distributed Computing |

| Advantages of R |

- | - Comprehensive statistical analysis |
- | - Open source, cross-platform |
- | - Collaborative community |

| Disadvantages of R |

- | - Inconsistent package quality |
- | - Memory management challenges |
- | - Slower compared to other languages |

| Applications of R |

- | - Data Science, Quantitative Analysis, Finance |
- | - Used by tech giants |

| R and Data Science |

- | - R and Python play major roles |
- | - Data science involves ML, statistics, data extraction, pattern recognition |
- | - Data analysis, finding patterns, and problem-solving |

| Tools for Data Science |

- | - R, Python, SQL, SAS, Tableau, MATLAB, etc. |
- | - R and Python are most used |

| R vs Python in Data Science |

- | - R: Statistical analysis, advanced techniques |
- | - Python: General purpose, web development, deep learning |

- | - Both good for finding outliers |
- | - Python better for web services |

...

Please note that this text representation is a simplified version of your original content, and the formatting may not perfectly capture the diagram's layout. If you need a more visually accurate representation, consider using a diagramming tool or software to create the actual diagram.

SIMULATED QA

Sure, here's a simulated Q&A based on the provided content:

Q1: What is Computational Data Analytics?

A1: Computational Data Analytics is a field that allows students to specialize in data science, encompassing areas like machine learning, deep learning, natural language processing, AI, visualization, databases, and high-performance computing. It builds on interdisciplinary core curriculum.

Q2: What are some examples of computational thinking?

A2: Computational thinking involves tasks like developing a chess strategy, creating and reading maps, and organizing complex to-do lists into manageable daily tasks.

Q3: What are the steps of Computational Thinking?

A3: Computational Thinking involves Abstraction (problem formulation), Automation (solution expression), and Analysis (solution execution and evaluation).

Q4: What are the principles of Computational Thinking?

A4: Computational Thinking encompasses decomposition, pattern recognition, abstraction, and algorithms. It's a problem-solving technique that can be practiced even before computer programming.

Q5: How does computational thinking benefit real-world problem-solving?

A5: Computational thinking helps break down large problems into simpler steps, aiding problem-solving in various domains, such as solving math problems or writing book reports.

Q6: What are computational skills?

A6: Computational skills involve quickly and accurately performing basic arithmetic operations like addition, subtraction, multiplication, and division using mental methods, calculators, or other tools.

Q7: How are models of computation classified?

A7: Models of computation are categorized into sequential models, functional models, and concurrent models.

Q8: What is the purpose of Computational Analytics?

A8: Computational models in analytics gather, filter, analyze, and present health information to guide doctors in disease treatment based on detailed patient characteristics.

Q9: How does Computational Data Science combine different fields?

A9: Computational Data Science combines elements of statistics, computer science, mathematics, and machine learning to identify trends, make predictions, and solve complex problems.

Q10: What is Data Analytics (DA)?

A10: Data Analytics involves examining data sets to identify trends and draw conclusions about the information they contain.

Q11: What is R programming used for?

A11: R is an open-source programming language widely used for statistical software, data analysis, and machine learning.

Q12: What are some advantages of R programming?

A12: R is comprehensive for statistical analysis, open-source, cross-platform, and has a growing user community.

Q13: How is R used in data analysis?

A13: R offers tools for basic statistics, static graphics, probability distributions, and data analysis.

Q14: What are the advantages of Python in data science?

A14: Python is versatile, emphasizing code readability, and is suitable for web development and deep learning.

Q15: How does Python compare to R for data analysis?

A15: R covers advanced techniques, while Python is better for web services and more common techniques. Both are good for finding outliers.

Q16: What are the main functionalities of R and Python for data science?

A16: R has inbuilt data analysis functionalities, while Python relies on packages like Numpy and Pandas.

Q17: Which programming language is better for deep learning?

A17: Python is better for deep learning with libraries like Keras, TensorFlow, and others.

Q18: How does R fare in terms of data visualization?

A18: R has rich data visualization tools through packages like ggplot2, while Python is catching up with packages like Matplotlib and Bokeh.

Q19: What are some applications of R?

A19: R is used in data science, quantitative analysis, finance, and is utilized by tech giants like Google, Facebook, and others.

Q20: How does R compare to Python in data science?

A20: Both R and Python play significant roles in data science, each with its own strengths and specialties. The choice between them depends on specific requirements.

QUIZ MCQ

Here are some multiple-choice questions (MCQs) based on the provided content:

Section: Computational Data Analytics

Q1: What is the primary focus of Computational Data Analytics?

- a) Social networking
- b) Artificial intelligence
- c) Data science specialization
- d) Computer hardware development

Q2: Which of the following is an example of computational thinking?

- a) Solving complex equations
- b) Painting a landscape
- c) Writing a novel
- d) Baking a cake

Q3: What are the steps of Computational Thinking?

- a) Visualization, Communication, Collaboration
- b) Abstraction, Automation, Analysis
- c) Calculation, Manipulation, Synthesis
- d) Identification, Experimentation, Conclusion

Q4: What are the principles of Computational Thinking?

- a) Explanation, Experimentation, Evaluation
- b) Decomposition, Pattern recognition, Abstraction, Algorithms
- c) Hypothesis, Observation, Inference
- d) Reflection, Refraction, Diffraction

Q5: What is the biggest benefit of computational thinking?

- a) Faster computer processing
- b) Improved typing skills
- c) Enhanced real-world problem-solving
- d) Advanced data visualization techniques

Section: R Programming

Q6: Which of the following best describes R programming?

- a) A video editing software
- b) An open-source programming language for data analysis
- c) A virtual reality gaming platform

d) A cloud computing service

Q7: What is the primary advantage of R programming language?

- a) Compatibility only with Windows operating system
- b) Closed-source nature
- c) Cross-platform support
- d) Limited functionality for statistical analysis

Q8: Which feature of R programming makes it a valuable tool for data analysis?

- a) Real-time 3D modeling
- b) Integration with social media platforms
- c) Inbuilt functionalities for data analysis
- d) Built-in video editing capabilities

Q9: What is a significant disadvantage of R programming language?

- a) Slow performance compared to other languages
- b) Lack of any packages or libraries
- c) Incompatibility with Windows OS
- d) Restricted to basic arithmetic calculations

Q10: Which major tech companies are using R programming?

- a) Amazon and Netflix
- b) Apple and Microsoft
- c) Google and Facebook
- d) Tesla and SpaceX

Section: R vs Python in Data Science

Q11: Which programming language is widely used for data science?

- a) Java
- b) Ruby
- c) C++
- d) Python

Q12: What are the specialties of R programming for data science?

- a) Covers advanced statistical techniques
- b) Specializes in web development
- c) Focuses on mobile app development
- d) Specializes in hardware design

Q13: Which programming language is better suited for deep learning?

- a) R
- b) Python
- c) Java

d) C#

Q14: What is one key domain where Python outperforms R in data science?

- a) Data visualization
- b) Machine learning
- c) Statistical analysis
- d) Database management

Q15: What is the primary advantage of Python over R in terms of packages?

- a) Python has more built-in functionalities for data analysis
- b) R packages are easier to find and use
- c) Python relies on fewer main packages
- d) R has a larger user community

Feel free to use these questions for your simulated QA session.

Sure, here are the answers to the multiple-choice questions:

Section: Computational Data Analytics

Q1: c) Data science specialization

Q2: b) Painting a landscape

Q3: b) Abstraction, Automation, Analysis

Q4: b) Decomposition, Pattern recognition, Abstraction, Algorithms

Q5: c) Enhanced real-world problem-solving

Section: R Programming

Q6: b) An open-source programming language for data analysis

Q7: c) Cross-platform support

Q8: c) Inbuilt functionalities for data analysis

Q9: a) Slow performance compared to other languages

Q10: c) Google and Facebook

Section: R vs Python in Data Science

Q11: d) Python

Q12: a) Covers advanced statistical techniques

Q13: b) Python

Q14: b) Machine learning

Q15: c) Python relies on fewer main packages

Please note that these answers are based on the provided content. If you have any more questions or need further clarification, feel free to ask!

FLASHCARDS/KEY TERMS

Here are the flashcards summarizing the key points from the provided text:

Flashcard 1:

Front:

Title: Introduction to Computational Data Analytics

Back:

- Computational data analytics is a field that provides depth and specialization in data science, including ML, deep learning, AI, and more.
- Examples of computational thinking include chess strategy, map reading, and task organization.
- Computational thinking involves steps like Abstraction, Automation, and Analysis.

Flashcard 2:

Front:

Title: Principals of Computational Thinking

Back:

- Computational thinking includes decomposition, pattern recognition, abstraction, and algorithms.
- Students can practice computational thinking before learning programming.
- Benefits of computational thinking include improved problem-solving skills.

Flashcard 3:

Front:

Title: Types of Computation and Purpose

Back:

- Models of computation: sequential, functional, concurrent.
- Computational models gather, analyze, and present health information for disease treatment.
- Computational Analytics enables scientific discovery through identifying patterns and anomalies.

Flashcard 4:

Front:

Title: R Programming Language

Back:

- R is an open-source language for statistical software and data analysis.
- R is cross-platform, available on Windows, Linux, macOS.
- R is used for machine learning, statistics, and data analysis.
- R's community is growing, and it's popular in the Data Science job market.

Flashcard 5:

Front:

Title: Features of R Programming

Back:

- R offers basic statistics, static graphics, and handles probability distributions.
- It has a wide availability of packages and supports distributed computing.
- Programs in R are written in IDEs and run using commands like "R file_name.r".

Flashcard 6:

Front:

Title: Advantages and Disadvantages of R

Back:

- Advantages: Comprehensive statistical analysis, open-source, cross-platform, welcoming community.
- Disadvantages: Some package standards, memory management issues, slower compared to some languages.

Flashcard 7:

Front:

Title: Applications of R

Back:

- Used for Data Science, quantitative analysis, finance, and by tech giants like Google, Facebook, etc.

Flashcard 8:

Front:

Title: R vs Python in Data Science

Back:

- R and Python play a major role in data science.
- R packages cover advanced techniques; Python is better for web services.
- Python is better for deep learning; R has inbuilt functionalities for data analysis.

Feel free to use these flashcards to study and review the concepts presented in the text. If you have any more information to include or specific points to emphasize, you can create additional flashcards accordingly.

- **Read PPT once**