

# STAT 129: ANALYZING AND PROCESSING BIG DATA

---

## In Workflow

1. MATH Committee Chair (taylorlm@csus.edu)
2. MATH Chair (kelce@skymail.csus.edu)
3. NSM College Committee Chair (mikkel.jensen@csus.edu)
4. NSM Dean (datwyler@csus.edu)
5. Academic Services (catalog@csus.edu)
6. Senate Curriculum Subcommittee Chair (curriculum@csus.edu)
7. Dean of Undergraduate (james.german@csus.edu; celena.showers@csus.edu)
8. Dean of Graduate (cnewsome@skymail.csus.edu)
9. Catalog Editor (torsetj@csus.edu)
10. Registrar's Office (w lindsey@csus.edu)
11. PeopleSoft (PeopleSoft@csus.edu)

## Approval Path

1. Wed, 13 Oct 2021 00:38:37 GMT  
Lisa Taylor (taylorlm): Approved for MATH Committee Chair
2. Thu, 14 Oct 2021 17:25:34 GMT  
Kimberly Elce (kelce): Approved for MATH Chair
3. Wed, 20 Oct 2021 23:26:29 GMT  
Thomas Krabacher (tsk): Rollback to MATH Chair for NSM College Committee Chair
4. Thu, 21 Oct 2021 17:38:07 GMT  
Kimberly Elce (kelce): Approved for MATH Chair
5. Thu, 21 Oct 2021 23:06:29 GMT  
Mikkel Jensen (mikkel.jensen): Approved for NSM College Committee Chair
6. Fri, 22 Oct 2021 16:32:05 GMT  
Shannon Datwyler (datwyler): Approved for NSM Dean

## New Course Proposal

Date Submitted: Wed, 13 Oct 2021 00:29:14 GMT

## Viewing: STAT 129 : Analyzing and Processing Big Data

Last edit: Thu, 21 Oct 2021 17:36:24 GMT

Changes proposed by: Clark Fitzgerald (223005263)

### Contact(s):

Name (First Last)	Email	Phone 999-999-9999
Kimberly Elce	kelce@csus.edu	8-5641

### Catalog Title:

Analyzing and Processing Big Data

### Class Schedule Title:

Big Data

### Academic Group: (College)

NSM - Natural Sciences & Mathematics

### Academic Organization: (Department)

Mathematics & Statistics

### Will this course be offered through the College of Continuing Education (CCE)?

No

### Catalog Year Effective:

Fall 2022 (2022/2023 Catalog)

### Subject Area: (prefix)

STAT - Statistics

**Catalog Number: (course number)**

129

**Course ID: (For administrative use only.)**

203532

**Units:**

3

**Is the primary purpose of this change to update the term typically offered or the enforcement of prerequisites at registration?**

No

**In what term(s) will this course typically be offered?**

Fall, Spring

**Does this course require a room for its final exam?**

Yes, final exam requires a room

**Does this course replace an existing experimental course?**

Yes

**This course replaces the following experimental course:**

STAT 196K - Analyzing and Processing Big Data

**This course complies with the credit hour policy:**

Yes

**Justification for course proposal:**

Statisticians and data scientists need specific training to work with large, complex data sets. The proposed "Big Data" course complements our existing Statistical Computing course (STAT 128) by covering some topics, such as tabular data, at a deeper level. It also covers more advanced topics, such as remote computing. Students who take both courses will be well-prepared for practical data analysis, whether in a graduate program, or in an entry level data scientist position. The creation of this course is part of an overall effort to increase the breadth of offerings for those interested in statistics and data science.

**Course Description: (Not to exceed 80 words and language should conform to catalog copy.)**

Statistical analysis of large, complex data sets. Topics include memory efficient data processing, the split-apply-combine strategy, rewriting programs for scalability, handling complex data formats, and applications such as statistical learning, dimension reduction, and efficient data representation. Students will access data and run code on remote servers.

**Are one or more field trips required with this course?**

No

**Fee Course?**

No

**Is this course designated as Service Learning?**

No

**Does this course require safety training?**

No

**Does this course require personal protective equipment (PPE)?**

No

**Does this course have prerequisites?**

Yes

**Prerequisite:**

(STAT 1 or STAT 50) and (MATH 26A or MATH 30) and (STAT 128 or CSC 20), or consent of the instructor.

**Prerequisites Enforced at Registration?**

Yes

**Does this course have corequisites?**

No

**Graded:**

Letter

**Approval required for enrollment?**

No Approval Required

**Course Component(s) and Classification(s):**

Lecture

**Lecture Classification**

CS#04 - Lecture /Recitation (K-factor=1 WTU per unit)

**Lecture Units**

3

**Is this a paired course?**

No

**Is this course crosslisted?**

No

**Can this course be repeated for credit?**

No

**Can the course be taken for credit more than once during the same term?**

No

**Description of the Expected Learning Outcomes: Describe outcomes using the following format: "Students will be able to: 1), 2), etc."**

Students will be able to:

- 1) Develop complete statistical computer programs based on high level directions, using standard software packages. Their programs will be complete in the sense that they start with processing raw data, and finish by producing final summaries and results necessary for reports.
- 2) Summarize their approach and conclusions for a data analysis problem through technical written reports with appropriate graphics.
- 3) Apply standard statistical techniques suitable for data larger than memory, for example, the split-apply-combine strategy for grouped data, memory efficient streaming statistics, discretization, and dimension reduction through principal components analysis.
- 4) Identify, extract, and summarize elements of interest from complex data sets, including tabular, hierarchical, streaming, image, and text data.
- 5) Perform data analysis using remote machines, which may include databases, remote compute clusters, and cloud services.
- 6) Accelerate and scale data analysis programs by identifying and fixing performance bottlenecks.

**Assessment Strategies: A description of the assessment strategies (e.g., portfolios, examinations, performances, pre-and post-tests, conferences with students, student papers) which will be used by the instructor to determine the extent to which students have achieved the learning outcomes noted above.**

Homework and Projects will assess outcomes 1, 2, 3, 4, 5, 6

Examinations will assess outcomes 2, 3, 4, 5, 6

Comprehensive final will assess outcomes 2, 3, 4, 5, 6

**For whom is this course being developed?**

Majors in the Dept

Minors in the Dept

**Is this course required in a degree program (major, minor, graduate degree, certificate?)**

No

**Does the proposed change or addition cause a significant increase in the use of College or University resources (lab room, computer)?**

No

**Will there be any departments affected by this proposed course?**

No

**I/we as the author(s) of this course proposal agree to provide a new or updated accessibility checklist to the Dean's office prior to the semester when this course is taught utilizing the changes proposed here.**

I/we agree

## **University Learning Goals**

**Undergraduate Learning Goals:**

Competence in the disciplines

**Is this course required as part of a teaching credential program, a single subject, or multiple subject waiver program (e.g., Liberal Studies, Biology) or other school personnel preparation program (e.g., School of Nursing)?**

No

## **GE Course and GE Goal(s)**

**Is this a General Education (GE) course or is it being considered for GE?**

No

**Please attach any additional files not requested above:**

Stat 129 syllabus 2021 10-12 (final).pdf

**Reviewer Comments:**

**Thomas Krabacher (tsk) (Wed, 20 Oct 2021 23:26:29 GMT):** Rollback: The CRC has approved the STAT 129 pending re-submission with the following changes. See email from Mikkel Jensen for further details: STAT 129 -- Assessment Strategies need to be integrated with syllabus -- Add grading scale to syllabus -- Need to expand justification. Could use justification from experimental course. Also add certificate information to justification

Key: 14601